

UNIVERSIDADE FEDERAL DO PARANÁ - UFPR
PROGRAMA DE PÓS-GRADUAÇÃO EM MÉTODOS NUMÉRICOS EM
ENGENHARIA – PPGMNE

ALESSANDRA MEMARI PAVANELLI

ANÁLISE DAS MEDIDAS DE DESEMPENHO DO SERVIÇO DE ATENDIMENTO
MÓVEL DE URGÊNCIA E EMERGÊNCIA EM CURITIBA-PARANÁ UTILIZANDO
CENÁRIOS ALTERNATIVOS NA DETERMINAÇÃO DE DIVISÕES GEOGRÁFICAS
DE ATENDIMENTO

CURITIBA
2014

ALESSANDRA MEMARI PAVANELLI

ANÁLISE DAS MEDIDAS DE DESEMPENHO DO SERVIÇO DE ATENDIMENTO
MÓVEL DE URGÊNCIA E EMERGÊNCIA EM CURITIBA-PARANÁ UTILIZANDO
CENÁRIOS ALTERNATIVOS NA DETERMINAÇÃO DE DIVISÕES GEOGRÁFICAS
DE ATENDIMENTO

Tese apresentada ao Programa de Pós-Graduação em Métodos Numéricos em Engenharia, na Área de Concentração em Programação Matemática, dos setores de Ciências Exatas e de Tecnologia da Universidade Federal do Paraná, como requisito parcial à obtenção do grau de Doutor.

Orientadora: Profa. Dra. Deise Maria Bertholdi Costa

Coorientador: Prof. Dr. Paulo Henrique Siqueira

CURITIBA
2014

P337a

Pavanelli, Alessandra Memari

Análise das medidas de desempenho do serviço de atendimento móvel de urgência e emergência em Curitiba-Paraná utilizando cenários alternativos na determinação de divisões geográficas de atendimento [manuscrito] / Alessandra Memari Pavanelli. — Curitiba, 2014.

176f. : il. [algumas color.] ; 30 cm.

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Setor de Tecnologia, Programa de Pós-graduação em Métodos Numéricos em Engenharia, 2014.

TERMO DE APROVAÇÃO

ALESSANDRA MEMARI PAVANELLI

ANÁLISE DAS MEDIDAS DE DESEMPENHO DO SERVIÇO DE ATENDIMENTO
MÓVEL DE URGÊNCIA E EMERGÊNCIA EM CURITIBA-PARANÁ UTILIZANDO
CENÁRIOS ALTERNATIVOS NA DETERMINAÇÃO DE DIVISÕES GEOGRÁFICAS
DE ATENDIMENTO

Tese aprovada como requisito parcial para obtenção do grau de Doutor no Curso de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, pela seguinte banca examinadora:

Heine m/basta

Prof.^a Dr.^a Deise Maria Bertholdi Costa
(Orientadora) Membro do PPGMNE/UFPR

Oyster Creek Bunt.

Prof.^a Dr.^a Ângela Olandoski Barboza
Membro da UTEPR

Membro da UTEPR

Prof. Dr. Lauro Cesar Galvão

Membro da UTFPR

Prof. Dr. Luiz Fernando Nunes

Membro da UTFPR

Paul

Prof.^a Dr.^a Luzia Vidal de Souza

Membro do PPGMNE/UFPR

Curitiba, 15 de abril 2014.

Dedico este trabalho a Deus, a
Nossa Senhora Aparecida e
aos amores da minha vida:
Genival, Giovana
e Alberto.

AGRADECIMENTOS

Agradeço a DEUS por estar sempre presente em minha vida, amparando-me nos momentos difíceis, dando força para superar os obstáculos e sempre me mostrando o caminho certo a seguir. Se cheguei até aqui, foi por Sua graça e misericórdia; sem Ele, nada teria conseguido. Agradeço a Nossa Senhora Aparecida, sempre intercedendo junto a seu filho nos meus pedidos de auxílio.

Embora uma tese tenha, por essência, a finalidade acadêmica de ser um trabalho individual, existem contribuições inestimáveis que não podemos deixar de destacar. Contribuições tão importantes que talvez sem elas não conseguíssemos concluir nosso trabalho. Por essa razão, gostaria de expressar meus sinceros e profundos agradecimentos às pessoas que fizeram parte desta caminhada acadêmica.

Agradeço primeiramente à Profa. Dr^a. Deise Maria Bertholdi Costa, profissional de extrema competência. Orientou-me e ajudou-me constantemente durante todos esses anos, sempre com muita alegria, entusiasmo e paciência, firme nas suas convicções e meiga nas atitudes. Acreditou que eu seria capaz. Se cheguei até aqui, por vezes desacreditando que seria possível, com toda a certeza foi porque tive uma pessoa maravilhosa ao meu lado, sempre mostrando o melhor caminho a seguir. Pessoa extraordinária e generosa, um exemplo a ser seguido. Saiba que serei eternamente agradecida e que, todos os dias, vai estar presente em minhas orações. Muito obrigada, Professora!

Agradeço ao Professor Prof. Dr. Paulo Henrique Siqueira, meu coorientador, pelas ideias importantíssimas para a construção da minha tese. Competente e muito dedicado em seu trabalho, ressalto aqui o excelente professor; afinal, fui sua aluna no doutorado.

Agradeço profundamente aos professores Ângela, Luiz e Luzia. Suas preciosas correções no meu trabalho de qualificação muito engrandeceram a minha tese. Todas as correções e sugestões foram valiosíssimas para o aprimoramento do trabalho final.

Agradeço ao professor Lauro Galvão, que hoje participa deste momento acadêmico tão importante em minha vida. Tenho certeza de que suas contribuições serão de suma importância.

Agradeço a todos os meus professores do PPGMNE com quem tive contato durante todos esses anos. Cada um, com seu estilo particular, contribuiu de maneira significativa para o meu aprimoramento intelectual e pessoal.

Agradeço, na pessoa do Sr. Jair, a empresa que disponibilizou todos os dados necessários para a realização deste trabalho.

Agradeço, na pessoa do Cel. Ávila, ao Colégio Militar de Curitiba, sempre apoiando e incentivando o aperfeiçoamento dos profissionais que ali trabalham.

De modo muito especial, quero agradecer aos meus colegas da disciplina de Física do Colégio Militar de Curitiba – os professores Ronaldo, Bach e Roberto – pelo apoio em todos os momentos dessa caminhada, sempre muito solícitos aos meus pedidos de ajuda.

Agradeço ao Reginaldo, militar que trabalha no Batalhão da Polícia Militar (BPM) de Curitiba, por ajudar-me nas análises estatísticas.

Agradeço às colegas Ana Maria e Prila, professoras de Inglês do Colégio Militar de Curitiba, por ajudarem na redação do Abstract deste trabalho. Às minhas amigas Tereza, Maísa e Lorena, também do Colégio Militar, que prontamente me ajudaram nas revisões finais do texto.

Agradeço à Maristela Bandil, nossa sempre secretária do CESEC, pessoa alegre, otimista, competente e sempre muito prestativa. E ao Jair, atual secretário, muito atencioso e prestativo.

Agradeço à minha família, por sempre me apoiar e me incentivar. Meu pai (*in memoriam*), se estivesse aqui hoje, transbordaria de orgulho. À minha mãe, exemplo de coragem, determinação e fortaleza, sempre acreditando que eu conseguiria alcançar meus objetivos. Aos meus irmãos, pessoas importantes em minha vida e

que amo muito, em especial à minha irmã e amiga, Sandra, por contribuir de maneira tão importante nas revisões de texto deste trabalho.

E neste momento quero agradecer de forma muito especial aos meus filhos, Giovana e Alberto, filhos amados e muito queridos. Sempre pacientes com meus estudos, entendendo que todo o tempo que tive de tirar do convívio familiar foi para um “bem maior”. Filhos, obrigada pela paciência em todos esses anos de estudo. Sem vocês eu não teria conseguido. Contudo, sei que servi de exemplo para vocês; afinal, vocês também gostam de estudar e isso muito me alegra.

Por último, mas não menos importante, quero agradecer ao meu marido, Genival. Muito obrigada por todo o apoio em todos esses anos de caminhada. Sem você, com certeza eu não estaria neste momento concluindo este curso tão sonhado. Você foi fundamental para a realização desta conquista. Você foi o primeiro a me incentivar a fazer este doutorado. Agradeço a Deus por você existir em minha vida.

Agradeço a todos que de alguma forma contribuíram para esta conquista tão desejada. Um sonho realizado!

Mais uma vez, a todos os meus sinceros agradecimentos.

“... E nunca considerem seu estudo como uma obrigação, mas sim como uma oportunidade invejável de aprender, sobre a influência libertadora da beleza no domínio do espírito, para seu prazer pessoal e para o proveito da comunidade à qual pertencerá o seu trabalho futuro.”

Albert Einstein

RESUMO

O acesso ao serviço de saúde de boa qualidade se reflete numa sociedade desenvolvida e satisfeita. Nos Serviços de Atendimento Emergencial (SAE) móvel, a satisfação está diretamente relacionada ao menor tempo de resposta, ou seja, menor tempo transcorrido desde um chamado até a chegada de um veículo ao local de atendimento. Visando melhorias nesse tipo de serviço, o objetivo deste estudo foi o de criar uma nova metodologia de análise do SAE móvel. O presente trabalho utilizou o modelo hipercubo de filas juntamente com uma nova metodologia baseada no *Ant Colony Optimization* (ACO), que neste caso tem a finalidade de gerar átomos independentes. O modelo hipercubo de filas tem por essência fornecer medidas de desempenho para cada cenário sugerido pelo pesquisador. Com isso, no intuito de analisar as configurações desse sistema de atendimento médico, foram implementados cenários alternativos quanto a divisão geográfica da cidade bem como horários diversos de congestionamento ao longo do dia. Para cada cenário proposto, calcularam-se algumas medidas de desempenho tais como *workload* dos veículos, tempo de atendimento às vítimas, tempos de viagens dos veículos, fração de atendimento, dentre outras. Essas medidas de desempenho foram analisadas e comparadas entre si. Para ilustrar a aplicação do modelo proposto, foi implementado um estudo de caso de uma empresa de atendimento de urgência e emergência da cidade de Curitiba-PR. Os resultados mostraram que o modelo é adequado para este tipo de serviço. Constatou-se, ainda, que as mudanças propostas no algoritmo baseado no ACO apresentaram cenários cujos resultados nas medidas de desempenho mostraram significativas reduções nos tempos de resposta.

Palavras-chave: Modelo hipercubo de filas. Medidas de desempenho. Atendimento de urgência e emergência.

ABSTRACT

The access to good quality health service is reflected in a developed and satisfied society. In mobile urgency and emergency care services, satisfaction is directly related to lower response time, that is, less time elapsed since a call to the arrival of a vehicle at the service location. In order to improve this type of service, the objective of this study was to create a new methodology of analysis of mobile urgency and emergency care services. This paper used the hypercube priority queuing model together with a new methodology based on Ant Colony Optimization (ACO), which in this case is intended to generate independent atoms. The hypercube priority queuing model essentially supplies performance measurements for each scenario suggested by the researcher. Aiming to analyze the settings for this medical care system, alternative scenarios concerning the city geographical division as well as varied traffic jam schedules along the day were implemented. For each proposed scenario, some performance measurements such as vehicle workload, victim assistance time, travel times of vehicles, among others, were calculated. These performance measurements were analyzed and compared among themselves. In order to illustrate the application of the proposed model, a case study of a company of urgency and emergency care in the city of Curitiba-PR was implemented. The results showed that the model is suitable for this type of service. It was further observed that the proposed changes in the algorithm based on the ACO presented scenarios whose results in the performance measurements demonstrated significant reductions in response times.

Key-words: Hypercube queuing model. Performance measurements. Care and emergency system.

LISTA DE FIGURAS

FIGURA 2.1 - VEÍCULOS DE ATENDIMENTO DE URGÊNCIA E EMERGÊNCIA	35
FIGURA 2.2 - VEÍCULOS APARELHADOS COM UTI.....	35
FIGURA 2.3 - DIVISÃO DA CIDADE DE CURITIBA	38
FIGURA 2.4 - DIVISÃO DE CURITIBA (BPM)	39
FIGURA 2.5 - DIVISÃO ILUSTRATIVA DE CURITIBA DE ACORDO COM A META-HEURÍSTICA BASEADA NO ACO	40
FIGURA 2.6 - ATENDIMENTOS EMERGENCIAIS REALIZADOS PELA EMPRESA	42
FIGURA 2.7 - DISTRIBUIÇÃO DO NÚMERO DE CHAMADAS AO LONGO DE UM DIA	42
FIGURA 3.1 - EXEMPLO DE FILA COM SEUS COMPONENTES.....	56
FIGURA 3.2 - ESTRUTURAS DOS SISTEMAS DE FILAS.....	61
FIGURA 3.3 - TRANSIÇÕES DE ESTADO DO MODELO $M/M/m$, QUANDO $n < m$	63
FIGURA 3.4 - TRANSIÇÕES DE ESTADO DO MODELO $M/M/m$, QUANDO $n \geq m$	64
FIGURA 3.5 - TRANSIÇÃO DE ESTADOS DO MODELO $M/M/m$	66
FIGURA 3.6 - REPRESENTAÇÃO DO ESPAÇO DE ESTADOS PARA UM SISTEMA DE ATENDIMENTO ESPACIALMENTE DISTRIBUÍDO, COM TRÊS SERVIDORES.....	73
FIGURA 3.7 - TAXA DE TRANSIÇÃO ENTRE OS SERVIDORES.....	77
FIGURA 3.8 - PONTE BINÁRIA BASEADO EM GOSS <i>ET AL.</i> (1989).....	89
FIGURA 3.9 - UM CONJUNTO COM 4 FORMIGAS.....	96
FIGURA 3.10 - FORMIGAS SAINDO SIMULTANEAMENTE E ALEATORIAMENTE.....	97
FIGURA 3.11 - CALCULANDO AS PROBABILIDADES	97
FIGURA 3.12 - CALCULANDO AS PROBABILIDADES	97
FIGURA 3.13 - CALCULANDO AS PROBABILIDADES	98
FIGURA 3.14 - FINAL DA PRIMEIRA ITERAÇÃO	98

FIGURA 3.15 - FLUXOGRAMA REPRESENTANDO O PSEUDO ALGORITMO BASEADO NO ACO PARA MÚLTIPLOS CAIXEIROS VIAJANTE.....	99
FIGURA 4.1 - DECOMPOSIÇÃO DE UM PROCESSO DE POISSON DE PARÂMETRO $\lambda > 0$	107
FIGURA 4.2 - HISTOGRAMA DA AMOSTRA DE SETE DIAS ANALISADOS..	107
FIGURA 4.3 - HISTOGRAMA DA AMOSTRA E DISTRIBUIÇÃO.....	108
FIGURA 4.4 - DIAGRAMA DE FLUXO PARA O ESTADO VAZIO (000000000)..	117
FIGURA 5.1 - DIVISÃO GEOGRÁFICA PELO BPM.....	120
FIGURA 5.2 - <i>WORKLOAD</i> DOS VEÍCULOS.....	122
FIGURA 5.3 - TEMPOS MÉDIO DE RESPOSTA PARA CADA ÁTOMO.....	123
FIGURA 5.4 - COMPARAÇÃO DOS WORKLOADS NOS DOIS PERÍODOS DO DIA COM A MESMA DIVISÃO USUAL DO SISTEMA.....	131
FIGURA 5.5 - DIVISÃO GEOGRÁFICA BASEADA NO ACO.....	132
FIGURA 5.6 - DIVISÃO GEOGRÁFICA BASEADA NO ACO COM ÁTOMOS DIFERENTES.....	137

LISTA DE TABELAS

TABELA 2.1 - QUANTIDADE DE ATENDIMENTOS POR TIPOS E DIAS DA SEMANA.....	33
TABELA 2.2 - ANÁLISE DO PERÍODO DE PICO (MANHÃ)	43
TABELA 3.1 - CORRESPONDÊNCIA ENTRE A NATUREZA E O ACO	90
TABELA 4.1 - PROPORÇÃO DE CHAMADOS EM CADA ÁTOMO DO SISTEMA DO CENÁRIO 1.....	106
TABELA 4.2 - INTERVALOS MÉDIOS ENTRE AS CHAMADAS SUCESSIVAS NO SISTEMA NO PERÍODO DE PICO.....	108
TABELA 4.3 - TAXAS MÉDIAS DE CHEGADA DOS CHAMADOS PARA CADA ÁTOMO POR PERÍODO.....	109
TABELA 4.4 - TEMPOS MÉDIO DE VIAGEM EM MINUTOS ENTRE SUBÁTOMOS OBTIDOS A PARTIR DE DADOS DA EMPRESA	110
TABELA 4.5 - MATRIZ DE LOCALIZAÇÃO DE VEÍCULOS	111
TABELA 4.6 - MATRIZ PREFERÊNCIA DE DESPACHO.....	112
TABELA 4.7 - TEMPOS E TAXAS MÉDIAS PARA CADA VEÍCULO	113
TABELA 4.8 - RELAÇÃO DO TEMPO DE VIAGEM COM O TEMPO DE ATENDIMENTO.....	114
TABELA 5.1 - <i>WORKLOAD</i> DOS VEÍCULOS E OS TEMPOS DE RESPOSTA	121
TABELA 5.2 - TEMPOS MÉDIO DE RESPOSTA DE CADA ÁTOMO	122
TABELA 5.3 - FRAÇÃO DOS DESPACHOS DE CADA VEÍCULO QUE SÃO INTERÁREAS DE COBERTURA.....	124
TABELA 5.4 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO.....	124
TABELA 5.5 - TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO.....	125
TABELA 5.6 - <i>WORKLOAD</i> DOS VEÍCULOS E OS TEMPOS DE RESPOSTA.....	126
TABELA 5.7 - TEMPOS MÉDIOS DE RESPOSTA DE CADA ÁTOMO	127

TABELA 5.8 - FRAÇÃO DOS DESPACHOS DE CADA VEÍCULO QUE SÃO INTERÁREAS DE COBERTURA.....	128
TABELA 5.9 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO.....	128
TABELA 5.10 - TEMPOS MÉDIOS DE VIAGEM DE CADA VEÍCULO PARA CADA ÁTOMO.....	129
TABELA 5.11 - COMPARAÇÃO <i>WORKLOAD</i> NOS DOIS PERÍODOS	130
TABELA 5.12 - <i>WORKLOAD</i> DOS VEÍCULOS E OS TEMPOS DE RESPOSTA	133
TABELA 5.13 - TEMPOS MÉDIOS DE RESPOSTA DE CADA ÁTOMO.....	134
TABELA 5.14 - FRAÇÃO DOS DESPACHOS DOS VEÍCULOS QUE SÃO INTERÁREAS	134
TABELA 5.15 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO.....	135
TABELA 5.16 - TEMPOS MÉDIOS DE VIAGEM DE CADA VEÍCULO PARA CADA ÁTOMO.....	136
TABELA 5.17 - <i>WORKLOAD</i> DOS VEÍCULOS E OS TEMPOS DE RESPOSTA	138
TABELA 5.18 - TEMPOS MÉDIOS DE RESPOSTA DE CADA ÁTOMO.....	138
TABELA 5.19 - FRAÇÃO DOS DESPACHOS DOS VEÍCULOS QUE SÃO INTERÁREAS.....	139
TABELA 5.20 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO.....	139
TABELA 5.21 - TEMPOS MÉDIOS DE VIAGEM DE CADA VEÍCULO PARA CADA ÁTOMO.....	140
TABELA 5.22 - COMPARAÇÃO DAS <i>WORKLOADS</i> NAS TRÊS DIVISÕES GEOGRÁFICAS.....	141
TABELA 5.23 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS.....	142
TABELA 5.24 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS.....	142
TABELA 5.25 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS.....	143
TABELA 5.26 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS NAS DUAS DIVISÕES GEOGRÁFICAS.....	144
TABELA 5.27 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS NAS DUAS DIVISÕES GEOGRÁFICAS.....	144

TABELA 5.28 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS NAS DUAS DIVISÕES GEOGRÁFICAS.....	145
TABELA 5.29 - COMPARAÇÃO DOS BACKUPS NOS DOIS CENÁRIOS..	146
TABELA 5.30 - COMPARAÇÃO NA FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO	146
TABELA 5.31 - COMPARAÇÃO DOS TEMPOS DE ESPERA NA FILA E TEMPO MÉDIO GLOBAL NOS TRÊS CENÁRIOS	147
TABELA 5.32 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO.....	148
TABELA 5.33 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO.....	148
TABELA 5.34 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO.....	149
TABELA 5.35 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO NOS TRÊS CENÁRIOS PROPOSTOS	149

LISTA DE ALGORITMO E GRÁFICO

ALGORITMO 3.1 - PSEUDO-CÓDIGO DO ALGORITMO ACO.....	94
GRÁFICO 4.1 - MÉDIA DO HORÁRIO DE CHEGADA DOS CHAMADOS	103

LISTA DE SÍMBOLOS E SIGLAS

B	representação vetorial de um estado do sistema, $B = \{b_1, b_2, \dots, b_n\}$;
b_i	representa o estado do servidor i;
E_{ij}	conjunto dos estados nos quais o servidor i é o primeiro servidor disponível na lista de despacho do átomo j;
EMS	<i>Emergency Medical Systems</i> ;
SAE	Sistema de Atendimento Emergencial;
SAMU	Serviço de Atendimento Móvel de Urgência;
$f_{ij}^{[1]}$	fração de despachos do servidor i ao átomo j que não implicam em tempo de espera na fila;
$f_{ij}^{[2]}$	fração de despachos do servidor i ao átomo j que implicam em tempo de espera na fila;
L_{ij}	matriz de localização do servidor i no átomo j;
VSG	veículo de serviço grave;
VSL	veículo de serviço leve;
Va	veículo que atende preferencialmente átomo a;
Vb	veículo que atende preferencialmente átomo b;
λ	taxa total de chegadas no sistema;
μ	taxa total de atendimento no sistema;
N	número de servidores no sistema;
N_A	número de átomos no sistema;
P	número de classes de usuários no sistema;
P_B	probabilidade de equilíbrio do estado B;
P_Q	probabilidade de saturação do sistema menos a probabilidade de todos os servidores estarem ocupados;
P_S	probabilidade de saturação do sistema;
ρ	carga média de trabalho no sistema;
ρ_i	<i>workload</i> - carga média de trabalho do servidor i;
\bar{T}	tempo médio de viagem no sistema;
\bar{T}_j	tempo médio de viagem ao átomo j;
t_{ij}	matriz dos tempos médios de viagem do servidor i ao átomo j;

\bar{T}_Q tempo médio de viagem para chamados na fila;

ONU Organização das Nações Unidas;

IBGE Instituto Brasileiro de Geografia e Estatística

SUMÁRIO

1 INTRODUÇÃO.....	21
1.1 OBJETIVOS DO TRABALHO.....	25
1.1.1 Objetivo Geral.....	25
1.1.2 Objetivos Específicos.....	25
1.2 JUSTIFICATIVA DO TRABALHO.....	26
1.3 ESTRUTURA DO TRABALHO.....	26
2 DESCRIÇÃO DO PROBLEMA E REVISÃO DE LITERATURA.....	28
2.1 HISTÓRICO DO SAE NO BRASIL.....	28
2.2 ESTRUTURA DO SAE.....	30
2.3 TIPOS DE SERVIÇOS DE ATENDIMENTO.....	31
2.4 ESTUDO DE CASO.....	32
2.4.1 Divisão de Curitiba e região metropolitana.....	36
2.4.1.1 Cenário 1.....	38
2.4.1.2 Cenário 2.....	40
2.5 OBTENÇÃO DOS DADOS.....	41
2.6 REVISÃO DE LITERATURA	43
2.6.1 Trabalhos relacionados a teoria de filas espacialmente distribuídos.....	44
2.6.2 Trabalhos relacionados a meta-heurística <i>Ant Colony Optimization</i> (ACO) e ao Problema do Caixeiro Viajante (PCV).....	48
3 MATERIAL E MÉTODOS.....	54
3.1 TEORIA DAS FILAS.....	54
3.1.1 Histórico.....	55
3.1.2 Definições de Teoria de Filas.....	56
3.1.3 Funcionamento do sistema.....	58
3.1.3.1 Forma com que os funcionários atendem os clientes.....	58
3.1.3.2 Forma de chegada dos clientes.....	58
3.1.3.3 A disciplina da fila.....	59
3.1.3.4 A estrutura do sistema.....	61
3.1.4 Modelo $M/M/1$ e $M/M/m$	62

3.1.5 Transição de estado.....	63
3.2 RELAÇÃO ENTRE A DISTRIBUIÇÃO DE POISSON E A DISTRIBUIÇÃO EXPONENCIAL.....	67
3.2.1 Distribuição de Poisson.....	67
3.2.2 Distribuição Exponencial.....	69
3.3 MODELO HIPERCUBO DE FILAS.....	71
3.3.1 Hipóteses do Modelo Hipercubo de Filas.....	73
3.3.2 Transições entre estados e equações de equilíbrio.....	76
3.4 MEDIDAS DE DESEMPENHO DO SISTEMA.....	82
3.5 ALGORITMO <i>ANT COLONY OPTIMIZATION</i>	88
3.5.1 Problema do caixeiro viajante.....	91
3.5.2 ACO Aplicada ao Problema do Caixeiro Viajante.....	92
3.5.3 ACO para o problema de múltiplos caixeiros viajantes (PMCV).....	94
 4 MODELAGEM, VALIDAÇÃO E APLICAÇÃO DAS HIPÓTESES DO MODELO HIPERCUBO DE FILA.....	 100
4.1 SISTEMA DE ATENDIMENTO EMERGENCIAL MÓVEL EM CURITIBA	100
4.2 DESCRIÇÕES DOS ATENDIMENTOS.....	101
4.3 PERÍODOS DE PICO.....	102
4.4 VALIDAÇÃO DAS HIPÓTESES DO MODELO HIPERCUBO DE FILAS...	103
4.4.1 Divisão da área em N_A átomos geográficos.....	104
4.4.2 Processo de chegada dos chamados.....	105
4.4.2.1 Cálculo dos λ_{jk}	107
4.4.3 Tempo de viagem.....	109
4.4.4 Servidores.....	110
4.4.5 Localização dos servidores.....	110
4.4.6 Despacho dos servidores.....	111
4.4.7 Política de despacho dos servidores.....	111
4.4.8 Tempo de atendimento.....	112
4.4.9 Dependência de atendimento em relação ao tempo de viagem.....	113
4.5 TRANSIÇÕES ENTRE ESTADOS E EQUAÇÕES DE EQUILÍBRIO.....	114

5 IMPLEMENTAÇÃO E RESULTADOS DO MODELO HIPERCUBO DE FILAS COM DIVISÃO GEOGRÁFICA USUAL E COM DIVISÃO BASEADA NO ACO.....	119
5. 1 CENÁRIO 1: PERÍODO DA MANHÃ, DAS 10 ÀS 14 HORAS.....	119
5. 2 CENÁRIO 1: PERÍODO DA TARDE, DAS 18 ÀS 21 HORAS.....	126
5. 3 COMPARAÇÃO DAS MEDIDAS DE DESEMPENHO DOS CENÁRIOS PERÍODO DA MANHÃ E DA TARDE.....	129
5.4 CENÁRIO 2: DIVISÃO GEOGRÁFICA DOS ÁTOMOS ATRAVÉS DA META-HEURÍSTICA BASEADA NO ACO COM MESMO NÚMERO DE PONTOS EM CADA ÁTOMO, NO PERÍODO DAS 10 ÀS 14 HORAS.....	131
5.5 CENÁRIO 2.1: DIVISÃO GEOGRÁFICA DOS ÁTOMOS ATRAVÉS DA META-HEURÍSTICA BASEADA NO ACO COM ÁTOMOS COM NÚMERO DE PONTOS DIFERENTES, PERÍODO DAS 10 ÀS 14 HORAS.....	136
5.6 COMPARAÇÃO DAS MEDIDAS DE DESEMPENHO DOS CENÁRIOS 1, 2 E 2.1.....	140
6 CONCLUSÕES E PROPOSTAS DE MELHORAMENTO.....	151
6.1 CONCLUSÕES.....	152
6.2 PROPOSTAS DE MELHORAMENTO.....	154
REFERÊNCIAS.....	155
ANEXOS	162

1 INTRODUÇÃO

Um dos maiores problemas em muitos países é a falta de acesso à saúde de boa qualidade, seja em serviços públicos ou particulares. Segundo a ONU (Organização das Nações Unidas, 2013), o Brasil, dentre todos os países da América Latina, encontra-se com o maior índice de insatisfação no que diz respeito à qualidade e ao atendimento do serviço emergencial. Um serviço de saúde de baixa qualidade é uma das grandes causas da pobreza de um país (ONU, 2013). Acredita-se que ao melhorar o serviço de atendimento emergencial é possível que se tenha uma melhor qualidade de vida para toda a população.

No caso do Serviço de Atendimento Emergencial (SAE) móvel, tanto nas áreas urbanas como nas rurais, bem como nas rodovias, a satisfação do usuário está ligada à rapidez e eficiência do atendimento. À medida que a população aumenta, crescem também os desafios em manter e até mesmo em melhorar este atendimento. A previsão, segundo pesquisas do IBGE, é que o Brasil tenha até o ano de 2025 uma população em torno de 228 milhões de habitantes. Isso significa um aumento de mais de 15% em apenas 13 anos. Para atender a esta nova demanda é necessário uma melhor reorganização dos atendimentos emergenciais móveis a fim de se obter respostas mais rápidas e eficientes às chamadas dos usuários. Tudo isso com o único objetivo de aumentar as chances de salvar vidas.

Para ressaltar a importância do tempo nos atendimentos, pode-se citar como exemplo, o aparecimento súbito de problemas cardíacos. Nesse caso, o tempo médio de resposta ao usuário do sistema deverá ser o menor possível, pois está diretamente relacionado à sobrevivência do mesmo. As chances de sobrevivência desse paciente diminuem à razão de 5% por minuto de espera. Com isso faz-se necessária, como já citou-se, a rapidez no tempo de resposta e um adequado atendimento no próprio local.

É importante ressaltar que o problema, muitas vezes, não se encontra apenas na quantidade de servidores ou veículos disponíveis, mas sim no planejamento, na localização dos servidores, inclusive na organização dos mesmos. Neste contexto, uma das aplicações possíveis é o modelo hipercubo de filas, que fornece as medidas de desempenho necessárias para analisar e implementar melhorias nos serviços prestados.

Os serviços de atendimento emergenciais apresentam um alto grau de incertezas quanto ao chamado, localização, tipo de chamado e até mesmo o horário do chamado. Nesse tipo de serviço o tempo de resposta deverá ser o menor possível, pois quanto maior o grau de incerteza mais rápido deverá ser o atendimento. Logo, a eficiência desse serviço será medida pelo tempo de resposta. Com isso, faz-se necessário que os veículos encontrem-se disponíveis no momento da ocorrência de um chamado emergencial. Não é suficiente apenas colocar um número maior de veículos à disposição. É preciso que se faça um adequado dimensionamento do sistema com a frota disponível. Para otimizar esse serviço, é importante que se analise o desempenho atual do sistema e se proponha mudanças, mesmo que pequenas, mas que gerem melhores resultados.

Sendo assim, com a finalidade de analisar o SAE móvel de Curitiba, buscou-se os dados em uma empresa particular. Essa empresa chegou em Curitiba-PR no ano de 1992. Sua principal finalidade é realizar atendimentos às solicitações de socorro, fora do ambiente hospitalar, com o principal objetivo de reduzir qualquer tipo de sequelas decorrentes da falta de socorro precoce. Seu funcionamento é de 24 horas por dia, todos os dias da semana. Cada equipe que atende a um chamado é composta por um médico especializado, um enfermeiro e um motorista que também é socorrista. Essa equipe presta atendimento em toda Curitiba e região metropolitana, atendendo em todos os locais requisitados, vias públicas, residências e trabalho. A cidade de Curitiba é assistida também pelo SAMU (Serviços de Atendimento Médico de Urgência) e pelo Corpo de Bombeiros, todos prestando serviços emergenciais de atendimento móvel.

De maneira geral, os SAE móveis são caracterizados por incertezas, tornando o sistema ainda mais complexo. Com isso faz-se necessário planejamento e organização por parte dos gerentes administrativos em analisar medidas de desempenho, tanto do ponto de vista do gerente, como por exemplo, carga de trabalho dos servidores, como do ponto de vista do usuário do sistema. Entre os parâmetros a serem otimizados tem-se o tempo de resposta a um chamado.

Desse modo, observando-se as características básicas do SAE, a ferramenta mais adequada neste trabalho, a fim de analisar algumas medidas de desempenho, é o modelo hipercubo de filas, proposto inicialmente por Larson, em 1974. É uma ferramenta baseada em teoria de filas espacialmente distribuídas e tem gerado resultados adequados. Segundo alguns pesquisadores como Mendonça e

Morabito (2000), Chiyoshi e Galvão (2000), Costa (2003), Takeda (2004), Iannoni (2005), Luque (2007), Souza (2010), Souza *et al.* (2013), entre outros. Essa ferramenta foi utilizada inicialmente no patrulhamento policial nos Estados Unidos e estendeu-se para serviços emergenciais da saúde, bombeiros e também para serviços semi-emergenciais, como por exemplo, reparos em redes elétricas. O interesse de pesquisadores nos SAE urbano vem aumentando muito nos últimos anos (SIMPSON e HANCOCK, 2009). Na área da saúde podem ser citados os trabalhos de Chaiken (1978), Takeda (2000), Takeda *et al.* (2004), Takeda *et al.* (2007), Mendonça e Morabito (2000), Morabito e Mendonça (2001), Chiyoshi *et al.* (2000), Luque (2007), Souza e Morabito (2008) e Souza (2010).

O modelo hipercubo de filas é uma poderosa ferramenta matemática que tem por essência fornecer medidas de desempenho. É um modelo descritivo e não fornece a configuração ótima, pois não é um modelo de otimização. Porém, fornece condições de avaliar o sistema a cada mudança sugerida de possíveis cenários alternativos e comparar as medidas de desempenho. Esse modelo torna-se uma ferramenta ainda mais robusta quando aplicado em conjunto com uma ferramenta de otimização. Ainda são poucos os trabalhos apresentados na literatura utilizando essa junção. Entre eles, destacam-se Chiyoshi *et al.* (2003), Galvão *et al.* (2003) e Galvão e Morabito (2008).

Os serviços de atendimento móvel de urgência (SAMU) de algumas cidades brasileiras foram analisados através do modelo hipercubo de filas e este se mostrou adequado em todos os casos. Os SAMUs têm características diferentes entre as cidades, logo cada problema deverá ser modelado considerando-se as particularidades regionais. Um exemplo é o SAMU de Campinas, SP, que possui duas classes distintas de usuários, a citar: básicos e avançados. Todos os veículos ficam centralizados e trabalham com uma fila de espera de no máximo dez usuários. No trabalho realizado por Takeda, foram avaliados cenários alternativos com descentralização dos veículos. O modelo se mostrou adequado para avaliação do sistema (TAKEDA, 2000).

Luque (2007) realizou um estudo para o SAMU de São Paulo no qual verificou-se que o modelo hipercubo seria inviável pela quantidade de veículos, já que o modelo hipercubo trabalha com o número de equações igual a 2^N , onde N refere-se ao número de veículos de atendimento. Como, a cidade de São Paulo, é a maior do país e por isso necessita um grande número de veículos, o modelo

forneceu uma quantidade muito grande de equações, extrapolando assim o número de variáveis e inviabilizando a execução do modelo computacionalmente.

Outro trabalho foi realizado no SAMU de Ribeirão Preto, SP, por Souza (2010). Nessa cidade, o SAMU atende três classes de usuários distintos, a citar: avançado, moderado e leve. Os veículos destinados ao atendimento avançado não podem atender casos leves ou moderados, gerando dessa forma prioridade na fila. Nesse caso, o modelo hipercubo teve que atender a essas prioridades. Estudou-se também diferentes períodos de pico ao longo do dia: manhã, tarde e noite. O modelo foi aplicado para cada período. Além da aplicação do modelo hipercubo foram realizadas simulações a fim de comparar as medidas de desempenho obtidas. Os resultados foram satisfatórios e mostrou-se que o modelo atende adequadamente esse tipo de problema.

A cidade de Curitiba conta com uma população de aproximadamente 1,8 milhões de habitantes, em uma área de 434,967 km² e uma média mensal de atendimentos emergenciais que varia de 4470 a 5500 chamados. Os serviços emergenciais móveis são atendidos basicamente pelo SAMU, que opera de forma integrada com os hospitais públicos, o corpo de bombeiros, em casos emergenciais e a empresa que nos disponibilizou os dados.

Essa empresa atende três tipos de chamados divididos em três categorias: casos graves, atendidos por veículos com UTIs (tipo D); casos não tão graves, atendidos por veículos de remoção (tipo B); e casos leves e consultas, atendidos por veículos de unidades de apoio (carros). O maior número de chamados acontece nos casos menos graves, ocasionando por vezes a ocupação de um veículo no momento em que acontece um chamado grave. Com relação a fila de espera não tem limite estabelecido e os veículos ficam locados em quatro pontos distintos da região metropolitana.

Para que seja possível aplicar a metodologia proposta nessa pesquisa, é necessário que a região estudada seja passível de ser dividida. No caso, a cidade de Curitiba foi dividida em quatro átomos geográficos (quatro regiões) de duas formas distintas: a primeira, de acordo com a divisão do Batalhão da Polícia Militar (BPM), divisão usual do sistema e a segunda divisão, de acordo com o algoritmo proposto neste trabalho, baseado no *Ant Colony Optimization* (ACO)

A fim de obter uma nova divisão geográfica da cidade de Curitiba, foi proposta uma mudança no algoritmo baseado na meta-heurística ACO. A

combinação do modelo hipercubo com uma meta-heurística a fim de propor novas configurações, ou seja, novas divisões, caracteriza a inovação de pesquisa apresentada neste trabalho.

1.1 OBJETIVOS DO TRABALHO

1.1.1 Objetivo Geral:

Criar uma metodologia de avaliação do Sistema Atendimento Emergencial (SAE) móvel, através das análises das medidas de desempenho relevantes fornecidas pelo modelo hipercubo de filas, de duas formas distintas: A primeira através da divisão já existente na região e a segunda divisão, através da mudança proposta pelo novo algoritmo baseado no *Ant Colony Optimization* (ACO), tendo por base os dados de uma empresa de Curitiba, PR.

1.1.2 Objetivos específicos:

1. Realizar um estudo de caso numa empresa particular a fim de verificar a aplicabilidade do modelo hipercubo de filas através da validação das nove hipóteses do modelo;

2. Construir um novo algoritmo baseado na meta-heurística ACO com a finalidade de propor novas configurações para o atendimento, ou seja, novas divisões geográficas, que chamaremos de átomos geográficos. Na primeira divisão geográfica todos os átomos possuem o mesmo tamanho (mesmo número de pontos) e na segunda divisão os átomos possuem tamanhos diferentes.

3. Comparar os diferentes cenários, com divisões geográficas distintas, obtendo através do modelo hipercubo, diversas medidas de desempenho. Além da comparação entre as medidas de desempenho nos cenários fazer também a comparação com os dados obtidos pela empresa.

4. Analisar diferentes períodos de congestionamento durante o dia. Esses subperíodos podem apresentar características temporais e geográficas diferentes a serem consideradas. Podem ainda fornecer medidas de desempenho diferentes que servem de base para analisar de forma distinta cada período crítico do dia.

O trabalho visa analisar o desempenho do sistema atual de funcionamento do SAE móvel e comparar com os cenários propostos, dessa forma, espera-se o surgimento de contribuições importantes para o processo de tomada de decisões por parte do gerente do sistema.

1.2 JUSTIFICATIVA DO TRABALHO

A presente pesquisa visa analisar o SAE móvel, através do estudo de caso em uma empresa particular de Curitiba. A empresa possui um sistema de informatização bem atualizado. Seus veículos são equipados com GPS e comunicação direta com a central de atendimento. Com esses equipamentos é possível agilizar o processamento das chamadas. No entanto, não existem avaliações periódicas que indiquem o desempenho do sistema. Sabe-se, por exemplo, que se houver, um aumento no número de veículos, é possível melhorar o nível do serviço oferecido. Contudo, faz-se necessário quantificar essa melhora no serviço, por meio de índice de desempenho, o que pode não ser uma tarefa trivial.

Dessa forma, o modelo hipercubo de filas juntamente com o ACO se enquadra perfeitamente como instrumentos adequados ao problema. Esse modelo fornece importantes medidas de desempenho, o que contribui para uma análise de todo o SAE móvel no que diz respeito à tomada de decisões mais coerentes, reduzindo, assim, os tempos de resposta do sistema.

1.3 ESTRUTURA DO TRABALHO

Este trabalho está organizado da seguinte forma: além da introdução considerada como a primeira seção, conta com mais cinco seções.

Na segunda seção é apresentada a descrição detalhada do problema abordado neste trabalho e a forma da obtenção dos dados para a realização do mesmo. Ao final é feita uma revisão de literatura sobre tema aqui tratado, mostrando os principais modelos de localização descritivos e prescritivos, bem como a meta-heurística *Ant Colony Optimization* (ACO).

Na terceira seção são apresentados os conceitos que envolvem as técnicas empregadas nesse trabalho, que são: teoria de filas, tipos de filas, modelo hipercubo de filas, meta-heurística ACO e por fim apresentam-se as mudanças propostas num

algoritmo baseado no ACO com a finalidade de criar divisões geográficas mais adequadas ao problema.

Na quarta seção são apresentados os experimentos realizados com os métodos propostos, bem como a validação das hipóteses do modelo hipercubo de filas ao problema abordado na seção II.

Na seção cinco tem-se a análise dos resultados obtidos nos cenários propostos no trabalho.

Na seção seis são apresentadas as conclusões por meio da análise dos resultados da seção anterior, são descritas as novas metodologias propostas e por fim sugestões para futuros trabalhos.

2 DESCRIÇÃO DO PROBLEMA E REVISÃO DE LITERATURA

Nesta seção é apresentado um breve histórico sobre o Sistema de Atendimento Emergencial (SAE) móvel no Brasil, bem como a estrutura e tipos de atendimento de maneira geral. São relatados de maneira objetiva, a obtenção e o tratamento dos dados coletados junto a uma empresa de saúde na cidade de Curitiba, Paraná. Finalmente, são apresentados resumos de alguns trabalhos correlacionados ao tema aqui tratado, mostrando os principais modelos de localização descritivos e prescritivos, bem como a meta-heurística *Ant Colony Optimization* (ACO).

2.1 HISTÓRICO DO SAE NO BRASIL

No Brasil, o Ministério da Saúde foi instituído em 25 de julho de 1953 com a Lei nº 1.920, embora sua história já houvesse iniciado em 1808. O ministério limitava-se apenas a atender os problemas mais importantes da saúde pública e tratava-se da principal unidade administrativa de ação sanitária do governo. Em 25 de fevereiro de 1967, com a implantação da Reforma Administrativa Federal, o Ministério da Saúde passou a ser o responsável pela formulação e coordenação da Política Nacional de Saúde, cujas principais responsabilidades eram: atividade médica e paramédica, ação preventiva em geral, vigilância sanitária de fronteira, controle de drogas, entre outras. O Ministério da Saúde passou por diversas reformas ao longo desses anos, entre essas destacam-se a instituída pela Constituição Federal de 1988, relativa a obrigatoriedade do estado em garantir saúde a toda a população. Nesse momento criou-se o sistema único de saúde (SUS). Outro marco importante foi a aprovação da Lei Orgânica da Saúde pelo Congresso Nacional no ano de 1990.

Podemos constatar que as reformas visando melhorias no sistema de saúde sempre estiveram presentes, contudo, devem ser implementadas mais rapidamente e precisam ser dinâmicas a fim de acompanhar o aumento populacional que é exponencial.

Conforme mencionado anteriormente, a população do Brasil tem previsão de alcançar cerca de 228 milhões de habitantes até o ano de 2025. Juntamente com a população, crescem os números relativos ao trânsito no Brasil como se pode

observar no último anuário estatístico do Departamento Nacional de Trânsito (DENATRAN). Em 2009, o país registrou uma frota de 59,3 milhões de veículos e mais de 400 mil acidentes de trânsito com vítimas. Segundo dados estatísticos da empresa, que forneceu dados para esse trabalho, o êxito no salvamento das vítimas é reduzido entre 7% a 10% a cada minuto de atraso, 50% dos acidentados podem morrer subitamente na primeira hora, 30% podem vir a óbito nas quatro primeiras horas e 20% entre duas e cinco semanas após o acidente. Tais números justificam a necessidade de sistemas de atendimento emergenciais mais eficazes e eficientes.

Sabe-se que a qualidade de vida de uma população está diretamente relacionada ao acesso à saúde de boa qualidade. Apesar dos avanços, os serviços de saúde no Brasil tanto em áreas urbanas como os prestados nas rodovias, necessitam de novas propostas de hierarquização e regionalização para uma assistência de melhor qualidade, maior rapidez e custos menores. Buscando atender essas necessidades, foram criados os sistemas de atendimento emergencial das cidades e das rodovias, os quais têm como objetivo socorrer as vítimas, prestando os devidos atendimentos no menor tempo possível. Desta forma, o SAE móvel, não poderá admitir fila de espera, pois tem a função de socorrer as vítimas de acidente e transportá-las até o hospital mais próximo, o mais rápido possível. Contudo, devido a algumas incertezas características do SAE, como por exemplo, a origem do chamado, a disponibilidade do servidor, a distribuição espacial, dentre outras, tornam o sistema altamente complexo, permitindo apenas a análise de algumas medidas de desempenho.

Entre as diversas medidas de desempenho existentes nesse tipo de sistema, o tempo entre um chamado e a chegada do servidor ao local da emergência, conhecido como tempo de resposta, é a medida de desempenho considerada mais importante do ponto de vista do usuário. Outra medida de desempenho que deve ser considerada importante, sob o ponto de vista do gerente do sistema, é a carga de trabalho do servidor, que é o tempo que ele permanece ocupado. Para que o sistema ofereça um serviço eficiente, de boa qualidade e de baixo custo ao usuário, essas medidas de desempenho, aparentemente contraditórias, são na realidade correlacionadas e devem ser gerenciadas igualmente com os recursos disponíveis no SAE móvel, a fim de atingir o objetivo maior que é a satisfação do usuário do sistema.

A partir da década de 60, quando foram definidos os primeiros modelos matemáticos nesse assunto, o interesse em problemas de localização de instalações teve um aumento significativo. São muitos os trabalhos de Pesquisa Operacional encontrados na literatura que tratam dos problemas de localização probabilísticos dos serviços de emergência. Podemos citar os sistemas de atendimento médico, patrulhamento policial e combate a incêndios. Os trabalhos diferem, entre outros aspectos, por tratarem de dois tipos distintos de serviço: emergenciais e não emergenciais. O modelo hipercubo proposto em Larson (1974), baseado em teoria de filas espacialmente distribuídas, tem se mostrado muito eficaz para analisar todos os tipos de serviços de localização, incluindo os SAEs móvel (LARSON, 1971).

2.2 ESTRUTURA DO SAE

Para que o SAE seja considerado eficiente, existem vários fatores que devem ser analisados. Contudo, o tempo gasto para o veículo chegar ao local do chamado, que deve ser o menor possível, é o mais relevante. Com isso, apesar da aleatoriedade e da distribuição espacial das chamadas, faz-se necessário uma constante atualização por parte do sistema sobre o local da ocorrência do chamado, o tipo de ocorrência e a localização dos servidores. É importante observar também que, para cada tipo de ocorrência, deverá ser designado um tipo específico de veículo, com pessoas qualificadas para o atendimento, podendo ser desde um caso grave, necessitando de médicos especializados e veículos aparelhados com UTI, até um simples chamado de remoção.

Segundo Larson (1972), a política de despacho em SAEs poderá ser definida seguindo alguns parâmetros:

- Quantidade de servidores e suas categorias específicas, com suas respectivas áreas de atuação em cada dia da semana;
- O critério de cada servidor para atender um determinado chamado;
- Determinação da localização de cada servidor (atendente);
- Lista de preferência de despacho para cada servidor;
- Definição das situações em que deve haver realocação de servidores.

Podem existir situações onde o servidor mais próximo da ocorrência encontre-se ocupado no momento da solicitação. Nesse caso, a política de atendimento do SAE designa automaticamente ao local da ocorrência, outro

servidor. Essa cooperação entre os servidores é uma das características do sistema de atendimento, sendo esse segundo servidor enviado ao local de atendimento denominado de *backup* (IANNONI, 2005).

2.3 TIPOS DE SERVIÇOS DE ATENDIMENTO

Os serviços de atendimento básicos oferecidos à população dividem-se em três categorias, segundo Gonçalves *et al.* (1994):

- Serviços de rotina: São serviços previsíveis, regidos por padrões estatísticos bem definidos. Normalmente apresentam bom desempenho. Alguns exemplos: coleta de lixo, entregas de encomendas, correios, entregas de jornais.
- Serviços semi-emergenciais: Apresentam alguma incerteza, porém com baixa solicitação de serviços. Apresentam desempenho satisfatório. Como exemplos: distribuição de água e esgoto, reparos na rede elétrica, serviço de telefonia e internet.
- Serviços de emergências: Nesse tipo de atendimento existe um alto grau de incerteza e a eficiência é medida através do tempo de resposta a um chamado. Exemplos são: policiamento, corpo de bombeiros e serviço de atendimento médicos de urgência.

Ao serem analisados os tipos de serviços prestados, percebe-se que nos dois primeiros tipos de atendimento, caso exista pequenos atrasos no tempo de atendimento, não haverá grandes problemas para a população. No entanto, nos serviços emergenciais, atrasos de qualquer natureza podem ser cruciais, ou seja, pode ser a diferença entre vida ou a morte do usuário. Conclui-se que nos serviços emergenciais, melhor será o desempenho do sistema quanto mais rápido for a resposta. Contudo, a rapidez depende de vários fatores. Não basta apenas, por exemplo, aumentar o número de veículos e sim dimensionar melhor o número veículos e profissionais, pois, no momento de um chamado o veículo precisa estar a disposição para atendê-lo. Dessa forma é necessária uma análise adequada do sistema emergencial, buscando um melhor grau de atendimento e como consequência tempos ociosos para os veículos.

Verifica-se então, que dependendo da análise determinística realizada, as conclusões podem ser equivocadas, uma vez que nem sempre são incorporadas as

variabilidades reais dos processos, ou seja, para obter uma boa avaliação do desempenho do sistema emergencial, faz-se necessária a utilização de ferramentas probabilísticas, a fim de obter indicadores de desempenho que retratem a realidade de forma mais precisa possível.

Nos serviços emergências predominam algumas características necessárias e fundamentais, mas a rapidez no atendimento com certeza é a maior dentre todas. Também existem os questionamentos em todos os serviços emergenciais difíceis de serem respondidos sem uma análise probabilística, como por exemplo: Qual o número necessário de veículos disponíveis para os atendimentos? Qual a localização ideal desses veículos? Quanto ao dimensionamento da equipe que prestará o socorro, qual o número de componentes de uma equipe para realizar um atendimento?

Várias pesquisas vêm sendo desenvolvidas com o objetivo de avaliar o desempenho do sistema. Na maioria desses trabalhos, as ferramentas mais adequadas são as probabilísticas, pois fornecem bons indicadores de desempenho que retratam fidedignamente a realidade dos usuários.

2.4 ESTUDO DE CASO

Neste trabalho, a fim de construir uma metodologia de avaliação do SAE móvel, serão analisados dados dos serviços de atendimento de urgência e emergência prestados por uma empresa que iniciou suas atividades na América Latina no ano de 1984, na cidade de Rosário, Argentina. A empresa começou a operar em Curitiba-PR no ano de 1992, em 1996 passou a atuar também em Porto Alegre, RS. Sua finalidade é realizar atendimentos às solicitações de socorro fora do ambiente hospitalar, com o principal objetivo de reduzir problemas decorrentes da falta de socorro precoce. Seus sistemas operam com políticas de despacho pré-determinadas, como por exemplo, somente um veículo pode ser despachado para atender a um chamado emergencial. Devido a essas características primordiais, o modelo hipercubo de filas, pode ser uma ferramenta muito útil a fim de direcionar o planejamento de sistemas de saúde emergencial.

A empresa, além de disponibilizar os dados, elucidou o funcionamento do sistema, contribuindo de maneira significativa para o levantamento dos dados, bem como a interpretação dos resultados obtidos após a aplicação da ferramenta

utilizada nesse estudo de caso. Conforme já mencionado, o sistema possui uma informatização que agiliza o processamento das chamadas, mas não existem avaliações periódicas que indicam o desempenho do sistema.

Segundo IBGE (Instituto Brasileiro de Geografia e Estatística) a grande Curitiba apresenta uma população de aproximadamente 1.776.761 habitantes, em uma área de 434,967 km² e uma média mensal de atendimentos emergenciais de 4.470 chamados. A cidade conta com uma parceria entre os SAEs móveis: o SAMU - Curitiba, que opera de forma integrada com os hospitais públicos; o corpo de bombeiros, que presta serviços não apenas de combate a incêndios, mas sempre que necessário presta também os primeiros socorros às vítimas, pois possui veículos equipados para esses tipos de atendimentos; e a cidade conta também com a empresa privada que disponibilizou os dados utilizados no presente trabalho. Todas possuem basicamente as mesmas características, ou seja, a mesma estrutura de atendimento emergencial.

De acordo com os dados fornecidos pela empresa, constatou-se que não existe diferença significativa de demanda quanto ao número de chamados nos diferentes dias da semana, conforme se pode observar na TABELA 2.1, exceto no sábado, contudo o gerente do sistema explica que esse dia é considerado dia de lazer, ou seja, poucos veículos e pessoas circulando nas ruas, logo, menor o número de chamados. Essa tabela mostra a distribuição do número de atendimento por tipo e por dia da semana.

TABELA 2.1 - QUANTIDADE DE ATENDIMENTOS POR TIPOS E DIAS DA SEMANA

Dia da semana	Emergência	Urgência	Outros	Total
Segunda- feira	36	454	515	1005
Terça-feira	25	440	489	954
Quarta-feira	33	412	436	881
Quinta-feira	35	431	459	925
Sexta-feira	33	301	290	624
Sábado	18	226	264	508
Domingo	33	311	277	621
Total	213	2575	2730	5518

O sistema conta com uma central de atendimento que recebe os chamados. Os atendentes recebem treinamento da empresa e fazem algumas perguntas ao

usuário quanto ao seu problema para poder classificar inicialmente o tipo da emergência. Após a triagem inicial, a chamada é transferida para um médico regulador e este dará um parecer clínico mais preciso juntamente com as primeiras orientações sobre os procedimentos que deverão ser tomados. Neste momento, classifica-se o chamado quanto à sua gravidade em: grave; não tão grave; ou leve. O tempo de atendimento é computado desde o momento da chamada, tempo de preparação da equipe, o tempo de espera na fila, o tempo de viagem, o tempo de atendimento (em cena) e o tempo de retorno à base, denominado tempo de resposta.

A pesquisa de campo foi iniciada no ano de 2012. Na época o sistema operava com 19 veículos, divididos em três categorias a citar: UTIs (tipo D), remoção (tipo B) e unidades de apoio (carros), conforme visto na (FIGURA 2.1). Dentre todos os veículos, apenas oito veículos foram utilizados no trabalho, pois o estudo foi dedicado ao atendimento de serviços considerados emergenciais. Desses oito veículos, quatro são destinados aos casos graves e quatro para casos não tão graves. Os casos considerados leves, como as remoções e consultas não foram considerados inicialmente nesta pesquisa. A diferença entre os veículos emergenciais é apenas com relação aos equipamentos, ou seja, na parte interna do veículo, conforme mostra a (FIGURA 2.2). A formação da equipe é feita sempre da mesma forma. Todos os veículos, desde a ambulância para simples remoção de pacientes estáveis, as unidades de suporte básico à vida e, principalmente as UTIs móveis, totalmente equipadas para suporte avançado à vida, são comandadas por médicos com experiência em atendimento de emergência e urgência médica, enfermeiros e socorristas.

Os veículos destinados ao atendimento de casos graves, eventualmente poderão atender a outros chamados, não tão graves, desde que os veículos destinados ao atendimento de casos não tão graves estejam ocupados. O mesmo vale para os veículos destinados ao atendimento de casos não tão graves.



FIGURA 2.1 - VEÍCULOS DE ATENDIMENTO DE URGÊNCIA E EMERGÊNCIA



FIGURA 2.2 - VEÍCULOS APARELHADOS COM UTI

Observando-se os dados contidos na (TABELA 2.1), verifica-se que 50% dos chamados são considerados emergência e urgência. Destes, aproximadamente 8% são casos emergentes e 92% casos urgentes. Considerando-se a totalidade dos chamados, 96% classificam-se como casos de urgência e outros (leves). Dentre os casos considerados urgentes e outros, os sintomas são: cefaleia, tosse forte, edemas, feridas cortantes, dor abdominal, entre vários outros. Já os casos de emergência, acontecem em menor escala. No entanto, podem levar a óbito rapidamente se não forem atendidos com muita rapidez e com equipamentos adequados a fim de estabilizar as condições do paciente até a chegada a um hospital.

Segundo dados da empresa, as doenças cardiovasculares são o principal motivo de óbito no Brasil. De acordo com o Portal Brasil (2011), as doenças cardiovasculares são responsáveis por 29,4% de todas as mortes registradas no país em um ano. Isso significa que mais de 308 mil pessoas faleceram principalmente de Infarto e Acidente Vascular Cerebral (AVC). Estudos realizados no Instituto Dante Pazzanese de Cardiologia, São Paulo, apontaram que 60% dessas vítimas são homens, com idade média de 56 anos. Contudo a doença pode atingir pessoas de qualquer idade, sobretudo após 65 anos. Diante dessas estatísticas, o Brasil encontra-se entre os 10 países com maior índice de mortes por esse tipo de doença. Segundo Quintana (2011), a Hipertensão Arterial Sistêmica (HAS) faz parte do grupo de fatores de risco que representam o maior percentual de mortalidade por doenças como AVC e infarto agudo do miocárdio. Existem também as sequelas permanentes oriundas das doenças cardiovasculares, incapacitando uma em cada três vítimas. Quando os sintomas aparecem subitamente, as chances de sobrevivência de um paciente diminuem à razão de 5% por minuto de espera. Por isso a necessidade de rapidez no tempo de resposta e adequado atendimento no próprio local.

2.4.1 Divisão de Curitiba e região metropolitana

A fim de definir regiões distintas de atendimento, condição necessária para a aplicação do modelo hipercubo de filas, a grande Curitiba foi dividida em quatro átomos (regiões) de duas formas independentes. Inicialmente a divisão foi realizada segundo os Batalhões da Polícia Militar (BPM) de Curitiba, divisão usual no sistema.

São elas: 12º BPM, 13º BPM, 20º BPM, 23º BPM, sendo que o município de Araucária foi adicionado ao polígono do 23º BPM e o de São José dos Pinhais ao do 20º BPM. Na sequência a divisão ocorreu através de uma meta-heurística baseada no *Ant Colony Optimization* (ACO). Nesta segunda divisão, inicialmente todos os átomos tinham o mesmo número de pontos, posteriormente a divisão foi realizada com número de pontos diferentes para cada átomo. O ACO original foi criado por Dorigo (1992), inicialmente com o objetivo de ser aplicado ao problema do caixeiro viajante. Nas duas divisões geográficas, BPM e ACO, cada átomo foi particionado conforme características da empresa para atendimento de urgência e emergência, totalizando dessa forma oito subátomos. A (FIGURA 2.3) mostra os respectivos bairros que compõem a cidade de Curitiba facilitando na sequência a interpretação dos bairros que compõem cada átomo de acordo com cada divisão geográfica.

Criou-se dessa forma dois cenários distintos em relação a divisão geográfica da cidade de Curitiba, o primeiro cenário, com a divisão do BPM e o segundo cenário, com a divisão da cidade realizada através da meta-heurística baseada no ACO. Analisou-se também neste trabalho dois intervalos de tempo, um no período da manhã e outro no período da tarde.

O objetivo de criar cenários diferentes foi o de realizar uma comparação nas medidas de desempenho obtidas através do modelo hipercubo de filas, uma vez que para cada divisão geográfica o modelo fornece diferentes medidas de desempenho e com isso novas análises do SAE móvel podem ser realizadas.



FIGURA 2.3 - DIVISÃO DA CIDADE DE CURITIBA
 FONTE: IPPUC INSTITUTO DE PESQUISA E PLANEJAMENTO URBANO DE CURITIBA
 (2001)

2.4.1.1 Cenário 1

A divisão da região realizada de acordo com o BPM, será chamada de cenário 1. Nesta divisão, os veículos, quando disponíveis, ficam localizados em apenas três átomos. Neste caso excetuando o átomo 23º BPM (em amarelo), onde não há base de emergência. As bases encontram-se nos bairros Juvevê (átomo 20 -

azul), Novo Mundo (átomo 13 - Rosa) e São José dos Pinhais (átomo 20 - azul) e Batel e Mário Tourinho (átomo 12 - marrom). O sistema admite fila de espera de um chamado por veículo. A (FIGURA 2.4) mostra a divisão da grande Curitiba em seus quatro átomos (BPM) e os respectivos bairros que o compõem.

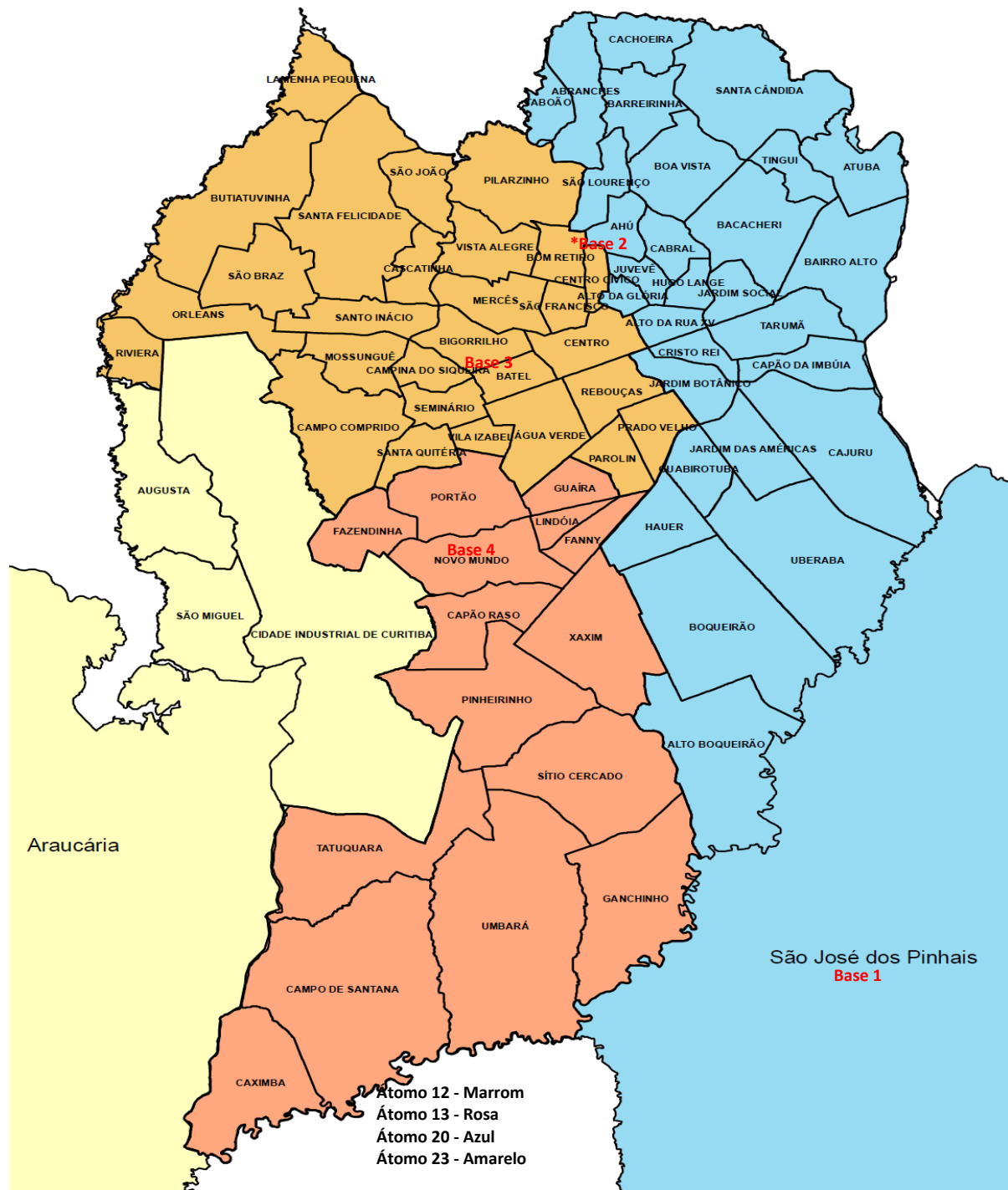


FIGURA 2. 4 - DIVISÃO DE CURITIBA (BPM)

Conforme mencionado anteriormente, dentro de cada átomo houve uma subdivisão devido à particularidade dos tipos de atendimento da empresa, classificados em casos graves (emergência) e não tão graves (urgência), os casos considerados leve, como as remoções e consultas, não foram considerados neste trabalho.

2.5 OBTENÇÃO DOS DADOS

Realizou-se a coleta de dados em dois momentos distintos. No primeiro momento analisou-se a quantidade e tipos de atendimentos ocorridos por mês ao longo do ano. Em seguida, coletou-se uma amostra de sete dias para validar as hipóteses do modelo hipercubo de filas. Essa amostra contém todas as informações dos atendimentos, para cada um dos veículos, tais como: horário das chamadas, local do atendimento, tempo de preparação da equipe, tempo de deslocamento na ida, tempo de atendimento, tempo de retorno à base, características da doença, veículo que prestou o atendimento. A escolha dessa amostra foi aleatória.

Também realizou-se várias entrevistas ao longo dos anos de 2012 e 2013 com o coordenador da empresa. Vale ressaltar, que não é fácil o acesso aos dados do SAE móvel. Contudo, depois de muito diálogo, a empresa disponibilizou dados confiáveis sob a condição de sigilo absoluto.

Segundo dados estatísticos da empresa verificou-se que não existe diferença significativa entre os meses com relação ao número de chamadas, exceto em meses de férias, conforme a (FIGURA 2.6). Constatou-se que existem diferenças no tipo e na quantidade dos chamados entre os dias úteis e nos finais de semanas, bem como em feriados prolongados, entretanto, essas diferenças, ainda segundo os dados estatísticos, permanecem dentro de um padrão de normalidade.

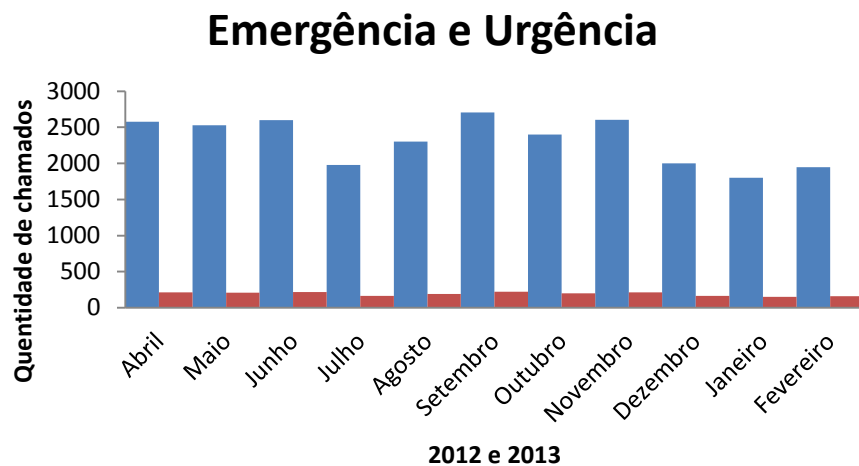


FIGURA 2.6 - ATENDIMENTOS EMERGENCIAIS REALIZADOS PELA EMPRESA

Quanto à segunda parte da coleta de dados, o objetivo foi o de analisar uma amostra aleatória de um determinado mês do ano, a fim de verificar se os chamados se comportam de maneira semelhante. O mês escolhido foi fevereiro do ano de 2013 e a amostra contém sete dias. Através da amostra, verificou-se que existem picos de chamada em determinados períodos do dia em todos os dias. Um desses períodos estava por volta das 10 horas e outro período em torno das 18 horas para cada dia, conforme apresentado na (FIGURA 2.7).

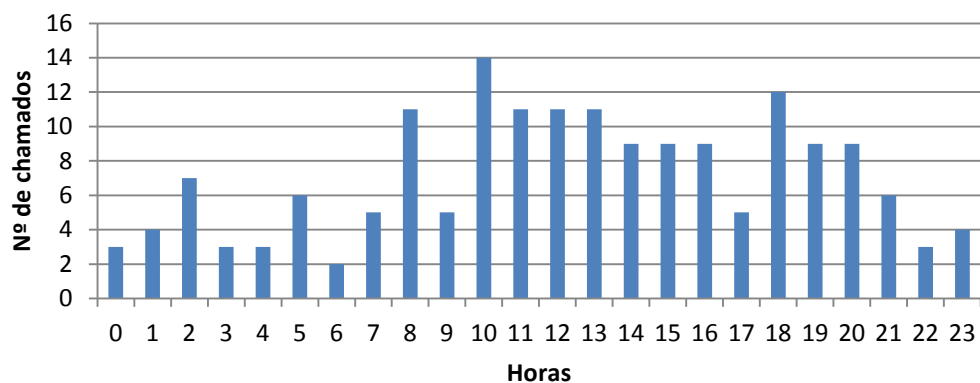


FIGURA 2.7 - DISTRIBUIÇÃO DO NÚMERO DE CHAMADAS AO LONGO DE UM DIA

Buscando encontrar o período de pico, ou seja, maior taxa média de chegada de chamados no sistema por intervalo de tempo, com o qual se pretendia trabalhar, dividiu-se inicialmente o dia em três turnos: manhã (das 9 às 16 horas), tarde (das 17 às 23 horas) e noite (0 às 8 horas). Posteriormente, dividiu-se cada

turno do dia em intervalos de horas, calculando a média e o desvio padrão para cada um destes intervalos. A (TABELA 2.2) apresenta apenas alguns intervalos de horas para o período da manhã, contudo, outros intervalos foram analisados.

Dentre os intervalos selecionados para compor a tabela, conforme pode-se observar, para o turno da manhã, o período de pico ficou caracterizado pelo horário compreendido entre às 10 e às 14 horas, uma vez que este período apresentou maior taxa de chegada dos chamados (número de chamados divididos pelo intervalo de tempo considerado), consequentemente a maior média, quando comparado com os outros intervalos. Todos os outros intervalos apresentaram médias inferiores.

TABELA 2.2 - ANÁLISE DO PERÍODO DE PICO (MANHÃ)

Períodos	1ºdia	2ºdia	3ºdia	4ºdia	5ºdia	6ºdia	7ºdia	Média	DP
9-14h	0,170	0,140	0,100	0,123	0,123	0,156	0,123	0,125	0,024
10-14h	0,180	0,130	0,120	0,142	0,123	0,150	0,121	0,131	0,022
11-15h	0,170	0,120	0,100	0,112	0,096	0,138	0,146	0,118	0,027
13-16h	0,130	0,100	0,140	0,122	0,116	0,133	0,127	0,128	0,013

A mesma análise foi realizada no turno da tarde e o período de pico ficou caracterizado pelo horário compreendido entre as 18 e 21 horas. O turno da noite apresentou a menor taxa de chegada de chamados. Com isso, nesse trabalho vamos considerar apenas o período de pico da manhã e tarde.

A amostra coletada conta com 1142 chamados, dos quais 3,32% não puderam ser georeferenciados por falta de dados no endereço. Trabalhou-se então com 1104 atendimentos no total, dos quais, segundo a divisão usual, 30,45% provem da região do 12º BPM, 22,5% da região do 13º BPM, 39,98% da região do 20º BPM e 7% da região do 23º BPM. A proporção dos chamados se manteve para o período de pico (manhã), onde ocorreram 239 chamados dos quais, 71 foram casos graves e o restante, 168 chamados, casos não tão graves, porém emergenciais.

2.6 REVISÃO DE LITERATURA

Listam-se a seguir trabalhos correlatos ao tema aqui apresentado, bem como a abordagem das técnicas com as quais se pretende trabalhar. Os trabalhos

dividem-se em modelos de teoria de filas, modelos de localização, modelo hipercubo de filas com otimização e meta-heurística *Ant Colony Optimization* (ACO) conhecido como colônia de formigas.

2.6.1 Trabalhos relacionados a teoria de filas espacialmente distribuídos

Vários modelos de localização probabilística foram criados com o objetivo de analisar e planejar os sistemas de atendimento emergencial, porém, os sistemas emergenciais possuem natureza estocástica. Devido a essas características, o modelo hipercubo de filas desenvolvido em 1974 por Larson, se enquadra nesse contexto, pois foi proposto para descrever sistemas onde servidores se deslocam para oferecer algum tipo de serviço (*server-to-customer service*). É baseado na teoria de filas espacialmente distribuídas. A aplicação original do modelo hipercubo foi para o problema de patrulhamento policial nos Estados Unidos. Diversos pesquisadores utilizaram-se dessa ferramenta em problemas de localização probabilística, dentre eles pode-se citar Chiyoshi *et al.* (2003), Galvão *et al.* (2003) e Galvão *et al.* (2005).

Em Batta *et al.* (1989) o modelo hipercubo de filas foi utilizado juntamente com uma heurística de substituição de vértice de máxima cobertura, visando determinar a localização dos servidores. Originalmente a heurística de substituição de vértices foi desenvolvida por Teitz Bart (1968) para o problema das p-medianas. O processo é considerado iterativo e utiliza o modelo hipercubo de filas com a finalidade de encontrar as probabilidades de ocupação de cada unidade de serviço. Mediante essas probabilidades de ocupação, calcula-se a cobertura máxima esperada. Esse procedimento é repetido a cada cenário sugerido, buscando sempre encontrar cenários que melhorem o valor da cobertura esperada.

Albino (1994) trata do dimensionamento de um sistema de atendimento às emergências geradas pelas interrupções no fornecimento de energia elétrica. As concessionárias de energia elétrica, em caso de interrupções de energia, necessitam atender a população no menor tempo possível. Esse tempo, que têm comportamento aleatório, é calculado pela soma do tempo de espera, tempo de deslocamento e tempo de execução do serviço. O modelo hipercubo de filas foi utilizado e mostrou-se adequado devido à complexidade geográfica e temporal do problema.

No trabalho de Takeda (2000), foi realizada uma análise no sistema de atendimento de urgência e emergência (SAMU) em Campinas, SP. O problema foi tratado com o modelo hipercubo de filas que se mostrou muito eficiente nesse tipo de abordagem, onde o serviço prestado é realizado com uma demanda aleatória e espacialmente distribuída. O modelo gera vários indicadores de desempenho para o sistema, podendo estes ser comparados com o sistema atual. Foram gerados cenários alternativos, novas localizações das ambulâncias, a fim de avaliar novas medidas de desempenho e verificar a melhor solução em termos de tempo de serviço prestado, já que o tempo é de suma importância para o atendimento emergencial.

No trabalho de Morabito e Mendonça (2001), o objetivo foi analisar o uso e a solução do modelo hipercubo para o caso de servidores não homogêneos (servidores com diferentes tempos médios de serviço). Sistemas com servidores não homogêneos podem ser encontrados em diversas situações reais, como por exemplo, nos Serviços de Atendimento Móvel de Urgência (SAMU) de algumas cidades brasileiras. A importância de se considerar explicitamente a não homogeneidade dos servidores no modelo hipercubo é mostrada inicialmente através de um exemplo ilustrativo. Em seguida demonstra-se que a solução para o caso não homogêneo pode ser obtida pelo método da eliminação de Gauss e o método de Gauss-Siedel. Os problemas teste para o método Gauss-Siedel foram gerados a partir de uma rede de 55 nós usada por Batta *et al* (1989), em modelos com 10 a 17 servidores, variando-se a taxa de ocupação do sistema de 0,1 a 0,9. Finalmente, para obter uma equação que permita estimar o número de iterações necessárias para resolver determinado problema realizou-se uma análise de regressão partindo de três variáveis: taxa de ocupação do sistema; número de servidores e uma variável binária de processamento das variáveis. Notou-se, através dos resultados obtidos, que o tempo para gerar os coeficientes do modelo hipercubo foi em média superior ao tempo para resolver as equações.

O trabalho de Costa (2003) foi realizado junto ao corpo de bombeiros na cidade de Curitiba, PR. O objetivo principal foi propor uma metodologia para determinação de zonas de atendimento para unidades de serviços emergenciais, estabelecendo assim novas posições de localização para as viaturas, de modo que suas áreas de atendimento fossem homogêneas de acordo com algum critério pré-estabelecido. Esse critério podia ser, por exemplo, tempo médio de espera para o

início de atendimento; tempo médio na fila de espera; entre outros. Foram coletadas algumas medidas de desempenho tais como tempo médio de deslocamento até o local do acidente, *workload* média das viaturas, entre outras, as quais foram utilizadas a fim de gerar uma estimativa entre o Atendimento Simulado e o Modelo Hipercubo. Foram estabelecidas, através da metodologia proposta, zonas de atendimento para as viaturas que diminuíram os tempos de deslocamento bem como os tempos em fila de espera.

Galvão *et al.* (2003) apresentaram um Problema de Localização de Máxima Disponibilidade (PLMD), definido por ReVelle e Hogan (1989), com a finalidade de alocar certo número de servidores de tal forma que atenda ao maior número de chamadas emergenciais possíveis, ou seja, ter sempre um servidor disponível em uma distância pré determinada. Esse trabalho foi uma extensão do PLMD desenvolvido por ReVelle e Hogan (1989), considerando o caso em que as taxas de ocupação foram calculadas individualmente. Trata-se de um serviço com taxas de atendimento diferenciadas entre os servidores. A ferramenta utilizada para tratar o problema foi o modelo hipercubo de filas, juntamente com uma heurística de substituição de vértices. A heurística foi testada em redes disponíveis na literatura: a rede de 55 vértices de Swain (1971) e as redes de 100 e 150 vértices de Galvão *et al.* (1996). Neste caso, a demanda associada a cada vértice dessas redes corresponde ao número de chamadas por dia originadas em cada área de demanda. Considerou-se também em alguns testes, que os tempos de atendimento entre os servidores fossem iguais. Os resultados obtidos com a heurística de substituição de vértices em geral confirmaram o comportamento esperado. Concluiu-se neste trabalho que as extensões propostas são viáveis neste tipo de problema.

Iannoni e Morabito (2006a, 2006b), utilizou em seu trabalho o modelo hipercubo de filas com algumas adaptações para o SAE em rodovias no estado de São Paulo. Devido às características básicas do SAE, o tempo de resposta a um chamado é, sem dúvida, uma das medidas de desempenho mais importantes do ponto de vista do usuário. Além da incerteza quanto aos chamados, existe também a incerteza quanto a localização dos servidores, fazendo-se necessário considerar os fatores probabilísticos na distribuição espacial e temporal dos chamados e dos servidores. No caso do SAE em rodovias, que apresenta serviços do tipo *server-to-costumer*, ou seja, o servidor se desloca para atender um determinado chamado, existem situações em que apenas algumas ambulâncias específicas do sistema

atendem chamados de uma determinada área, bem como casos onde duas ou mais ambulâncias podem atender a um único chamado. Dessa forma o modelo hipercubo de filas foi adaptado para situações de múltiplos despachos bem como a diferenciação entre os servidores. Foram propostos neste trabalho três modelos distintos de variações no modelo hipercubo, a fim de analisar configurações diferentes do SAE em rodovias. Concluiu-se que o modelo hipercubo de filas pode ser modificado, e os resultados obtidos mostram que os modelos são efetivos para avaliar as principais medidas de desempenho do SAE.

Iannoni *et al.* (2008a, 2008b), também utilizaram uma combinação do modelo hipercubo de filas com algoritmo genético. O objetivo foi o de otimizar a configuração e operação dos sistemas médicos emergenciais em rodovias. A junção dessas técnicas visa apoiar basicamente duas tomadas de decisões por parte da empresa: a primeira no que diz respeito à localização das bases dos veículos e a segunda quanto ao dimensionamento das regiões de cobertura de cada base da rodovia. Tudo isso, visando reduzir o tempo de resposta às vítimas. Foram utilizados dois estudos de caso realizados em rodovias brasileiras; o SAE dos Anjos do Asfalto, Morabito e Mendonça (2001) e o SAE Centro vias, Iannoni (2006), com o objetivo de aplicar e validar os métodos propostos. Concluiu-se que a junção do modelo hipercubo e filas com a meta-heurística algoritmo genético é promissora, pois forneceram medidas de desempenho que podem ser melhoradas realocando-se as bases dos veículos bem como variando o tamanho das áreas de cobertura. Os tempos computacionais foram satisfatórios para uma quantidade de veículos inferior a dez e inviável para uma quantidade de veículos superior a esse valor.

Morabito e Iannoni (2008) apresentam extensões do modelo hipercubo de filas com algoritmo genético a fim de otimizar a configuração e operação de sistemas de atendimento emergenciais em rodovias. O objetivo de tais procedimentos, inicialmente, é utilizar um método para localizar as bases de ambulância ao longo da rodovia e em seguida estender a abordagem para apoiar as duas decisões combinadas, ou seja, localizar as bases das ambulâncias e dimensionar as regiões de cobertura de cada base na rodovia. Foram analisados os resultados de dois estudos de caso em rodovias brasileiras.

Souza (2010) apresenta um estudo de caso no SAMU de Ribeirão Preto, SP, cujo objetivo foi o de aplicar o modelo hipercubo para considerar fila com prioridade e propor uma abordagem para múltiplas configurações de localização das

ambulâncias, explorando variações importantes da demanda e do serviço ao longo do dia. Diversos cenários alternativos foram analisados e os resultados foram muito interessantes, ajudando na tomada de decisão por parte dos operadores do sistema de atendimento da saúde.

2.6.2 Trabalhos relacionados a meta-heurística Ant Colony Optimization (ACO) e ao Problema do Caixeiro Viajante (PCV)

O algoritmo de otimização por colônia de formigas, oriundo do inglês, Ant Colony Optimization (ACO), é uma heurística baseada em probabilidade. Foi inicialmente estudada por Dorigo (1992), em um de seus trabalhos. É baseada no comportamento forrageiro das formigas ao saírem de seus ninhos em busca de alimento. O algoritmo busca uma solução para problemas computacionais que envolvam caminhos em grafos. Originalmente foi aplicado ao problema do caixeiro viajante (PCV).

Silva (2003) apresenta em seu trabalho uma completa avaliação do desempenho da meta-heurística *Ant Colony Optimization* (ACO) aplicada em grandes instâncias do problema de cobertura de conjuntos (*Set-Covering Problem* SCP). A tese procura analisar a influência dos parâmetros desta heurística sobre a qualidade dos resultados alcançados. Utiliza métodos estatísticos comparativos e de regressão, mas o principal método utilizado a fim de analisar o modelo é um rigoroso delineamento de um projeto experimental visando avaliar a adaptação da meta-heurística ACO em problemas de otimização combinatorial. Buscou-se estudar grandes instâncias devido a sua complexidade (NP-Completo), e dessa forma ajudar na resolução de problemas reais. Foram analisados, através de métodos experimentais, a meta-heurística ACO com adaptações para SCP, *Ant System* e MAX-MIN. Os resultados obtidos com as adaptação do ACO foram inferiores aos resultados obtidos pelo método busca local baseado no trabalho de Jacobs e Brusco (1995).

Baba *et al.* (2004a, 2004b) trata de problemas relacionados à programação e roteirização de veículos para o transportes de pessoas com deficiências e dificuldades de se locomover, na Empresa de Desenvolvimento Urbano e Social (URBES), localizada no município de Sorocaba-SP. É um problema relacionado à coleta e entrega com restrições de janela de tempo, capacidade do veículo e tempo

máximo de viagem. Conhecido na literatura como *dial-a-ride*. O objetivo nesse tipo de problema é construir uma programação de roteiros de modo a maximizar o número de solicitações atendidas e minimizar o custo de atendimento. O problema pode ser resolvido através dos métodos exatos ou heurísticos. Nesse trabalho utilizou-se uma heurística composta de duas fases, a primeira está relacionada com a inserção paralela proposta por Toth e Vigo (1997) e a segunda utilizando a meta-heurística colônia de formigas. Os resultados obtidos mostram-se promissores para o uso dessa meta-heurística.

Prestes (2006) apresenta em seu trabalho uma análise estatística de abordagens heurísticas aplicadas ao Problema do Caixeiro Viajante (PCV). O objetivo é avaliar o desempenho de cada abordagem em relação ao tempo computacional até a obtenção da solução ótima para uma determinada instância do PCV. Foi realizada uma análise experimental de abordagens heurísticas ao PCV. Para tanto foi utilizada uma metodologia chamada Análise de Sobrevivência (AS), auxiliada pelo teste de Log-Rank para o teste da hipótese da igualdade entre as funções de sobrevivência. Neste caso, analisou o tempo computacional necessário até a obtenção da solução ótima. Para melhor compreensão, as abordagens avaliadas foram divididas em três classes: Algoritmos Lin-Kernighan, Evolucionários e Otimização por Nuvens de Partículas. Na AS, apenas um evento foi registrado, quando a solução ótima foi encontrada. As soluções próximas do ótimo e o instante que esta solução foi obtida não foram registrados. Com isso, nas instâncias acima de 1500 cidades, muitas observações foram censuradas. Sugere-se considerar para AS, ao invés da solução ótima, soluções que tenham uma determinada taxa de afastamento (maior que zero) em relação ao ótimo.

Donati *et al.* (2008) tratam do problema de roteamento de veículos com dependência do tempo. Trata-se de encontrar a melhor rota no menor tempo. O tempo está relacionado com a velocidade que está interligada com os dias e horários da semana. Quando se trabalha com restrição de tempo, por exemplo: as janelas de tempo de entrega para o cliente, se faz necessário considerar o tráfego, pois, influenciará na velocidade do veículo. A otimização nesses casos consiste em minimizar o tempo de viagem e o número de veículos. As otimizações foram baseadas no algoritmo *Ant Colony System* (ACO) e nos procedimentos de busca local, uma vez que os métodos exatos despendem um elevado tempo computacional. O algoritmo ACO, muito utilizado no problema do caixeiro viajante e,

com excelentes resultados, foi adaptado para problemas com janela de tempo de roteamento juntamente com busca local, tornando-se uma metodologia robusta.

Carvalho e Yamakami (2008) propuseram uma hibridização da meta-heurística de otimização de colônia de formigas juntamente com algoritmos genéticos a fim de resolver o problema do caixeiro viajante. O PCV consiste na determinação de uma rota que inicia aleatoriamente em uma cidade i passando por todas as outras cidades j , de um conjunto de cidades, uma única vez e retornando a cidade inicial da rota, percorrendo uma distância mínima total. Essa hibridização visa melhorar a capacidade de busca do algoritmo, ou seja, melhorar a velocidade de convergência e a qualidade das soluções. A fim de avaliar o desempenho do algoritmo proposto, foram utilizados seis grafos completos com distâncias simétricas retirados de uma base de dados TSPLIB. Os resultados obtidos com o modelo proposto foram eficientes e promissores para a resolução do PCV. Todos os valores obtidos com a hibridização foram superiores quando comparados com o algoritmo colônia de formigas ou o algoritmo genético aplicado separadamente.

O trabalho de Yildirim *et al.* (2009) tratou de problemas de roteamento de veículos com janela de tempo com o objetivo principal de minimizar a distância total. Foi proposta uma mudança no algoritmo *Ant Colony Optimization* a fim de resolver o problema levando em consideração o tempo do trajeto. A incorporação do tempo de viagem foi através da função da velocidade, uma vez que existe uma forte relação com o tempo. Realizou-se uma completa avaliação sobre a teoria existente em relação ao assunto que trata da abordagem proposta. A otimização sobre uma rede de distribuição é de fundamental importância tanto na literatura como em aplicações do cotidiano. O Problema de Roteamento de Veículo (PRV) pode ser visto como uma generalização do Problema do Caixeiro Viajante (PCV), considerando apenas clientes no lugar das cidades e veículos no lugar de caixeiros. O roteamento de veículos consiste em determinar os roteiros para atender um conjunto de clientes situados em pontos distintos buscando minimizar o custo total. Cada ponto deve ser visitado uma única vez e a demanda de clientes atribuída a cada rota é previamente determinada e não poderá ser extrapolada. O principal objetivo é percorrer a menor distância, bem como reduzir o número de veículos. A dependência em relação ao tempo acontece de forma estocástica e determinística. Os resultados mostraram-se bem interessantes.

Pereira *et al.* (2010) utilizou a associação da meta-heurística colônia de formigas e a técnica de busca local a fim de obter melhores resultados para a segmentação de imagens. Segmentar significa subdividir uma imagem em pequenas partes para extrair a parte que julgar importante. Diversas são as técnicas para fazer essa segmentação, a manual baseada no reconhecimento visual, a automática simulando o comportamento de especialistas para reconhecer regiões homogêneas, baseando-se em suas propriedades espectrais e espaciais, entre outras. O objetivo desse trabalho é o de propor a meta-heurística Colônia de formigas juntamente com a técnica busca local, a fim de obter os melhores resultados. A técnica proposta foi aplicada em três imagens com os mesmos parâmetros α e β em todas elas. Para todas as imagens constatou-se que o processo de busca por ACO foi capaz de investigar soluções ainda não encontradas de modo a explorar da melhor maneira possível o espaço de soluções. Com isso diminuiu a possibilidade de cair em um ótimo local. A busca local foi aplicada ao final de cada solução construída na tentativa de melhorar o valor da função objetivo. Os resultados mostraram que as técnicas são muito eficientes nesse tipo de problema.

O trabalho de Silva (2012) visa desenvolver e validar o algoritmo de Formigas Paraconsistentes, que é um sistema híbrido da lógica Paraconsistente com a meta-heurística Colônia de Formigas. O algoritmo foi testado em problemas de otimização em sistemas elétricos de potência. Pode atuar tanto no espaço de busca discreto bem como no espaço de busca contínuo. Uma boa solução para os problemas onde os algoritmos exatos requerem altos gastos computacionais. A utilização da meta-heurística colônia de formigas foi pelo fato de ser uma técnica indicada para problemas de otimização combinatória, onde uma sequência de valores discretos devem ser obtidos para a solução do problema. Os resultados se mostram promissores para a resolução dos problemas de otimização.

Weyland *et al.* (2013), tratam do problema de roteamento de veículos para a coleta de óleo de cozinha em Bali - Indonésia. Os óleos são recolhidos a fim de serem reciclados e assim obter combustível biodiesel. Propõem-se para tanto, uma heurística para resolver esses problemas e apoiar na tomada de decisão quanto a localização ideal, bem como o número adequado de veículos que fazem a coleta. Existem 308 pontos de recolhimento e a quantidade de latas a serem recolhidas varia de 1 a 22 latas por ponto. O tempo de viagem deve estar dentro de um domínio razoável. Foi utilizado o algoritmo de busca local a fim de otimizar as rotas para os

diferentes veículos e dias. Os resultados obtidos foram muito bons no sentido de que foi possível obter as cargas de trabalho entre os veículos, e também obter informações sobre o número ideal de veículos, bem como as melhores localizações.

Neia *et al* (2013) apresentam em seu trabalho um exemplo ilustrativo que trata do Problema de Roteamento de Veículo Capacitado (PRVC), que é uma generalização do Problema do Caixeiro Viajante (PCV). O PCV é um problema simples de ser enunciado e compreendido, contudo muito difícil de ser resolvido até a otimalidade. Nesses estudos o problema é resolvido por duas meta-heurísticas evolucionárias, algoritmo genético e otimização por colônia de formigas (ACO), a fim de fazer uma comparação. A meta-heurística ACO inicialmente foi aplicada ao PCV em sua forma clássica, ou seja, apenas um veículo percorre todos os pontos, em seguida, o problema foi resolvido em sua forma original, ou seja, com cinco veículos percorrendo 31 pontos. A instância utilizada a fim de aplicar as técnicas é uma das instâncias de benchmark, Augerat *et al* 1998, que é referenciada por A-n32-k5. O Roteamento ou roteirização de veículos consiste em: determinar os roteiros para atender um conjunto de clientes situados em pontos distintos buscando minimizar o custo total; cada ponto seja visitado uma única vez e a quantidade transportada por veículo não exceda o limite da capacidade. A cada iteração, cada formiga (veículo) constrói uma solução para PRVC, movendo-se para o próximo nó. A lista tabu é usada para verificar quais clientes podem ser escolhidos. Isso se repete até a capacidade ser alcançada. Quando a capacidade máxima é alcançada o processo se repete para a próxima formiga, desconsiderando os clientes já visitados. Cada formiga é iniciada em um cliente diferente. A transição poderá ocorrer de duas formas, através da função da probabilidade ou com a regra pseudo aleatória proporcional. Quanto a atualização do feromônio, está pode ocorrer tanto no final da iteração ou durante o processo. O critério de parada pode ocorrer depois de um determinado número de iterações ou pela estagnação da solução. Os resultados obtidos mostraram que o ACO foi superior ao AG quando testado com apenas um veículo. Já com cinco veículos o AG mostrou-se mais eficiente. De qualquer forma, apesar das duas ferramentas não garantirem a otimalidade, ambas apresentaram boa qualidade levando em conta o esforço computacional. São procedimentos intuitivos que exploram as especificações do PRV de forma inteligente. Contudo os resultados poderão ser melhorados. Para tanto faz-se necessário a contínua pesquisa na ciência por métodos híbridos eficientes.

Mulati *et al* (2013) apresentam a fundamentação teórica e a aplicação da meta-heurística Otimização por Colônia de Formigas (ACO). O ACO é uma meta-heurística inspirada no comportamento forrageiro das formigas e que em seguida foi acrescido técnicas de busca local. Possui três características básicas: algoritmo construtivo guloso; comportamento auto catalítico e computação distribuída. Dentre várias aplicações do ACO em problemas de otimização, esses autores optaram por validar a técnica através de dois problemas distintos, o primeiro trata do Problema do Caixeiro Viajante (PCV) e o segundo do Problema de Cobertura de Conjuntos (PCC). No primeiro caso trata do relacionamento espacial do problema em analogia com o comportamento das formigas reais, como exemplo podemos citar o agendamento de uma máquina de fazer furos em placas de circuito. Onde os furos são as cidades e os custos da viagem é o tempo que se leva para mover a cabeça da perfuração de um orifício para o próximo. O segundo caso (PCC), explora um problema de otimização combinatória puramente matemático, aqui pode-se citar o escalonamento de pessoal em trânsito urbano ou mesmo escalonamento de pessoal em aeronave. O trabalho procurou analisar resultados sem se prender ao tempo de processamento nem a comparações rígidas as diferentes abordagens. O ACO foi comparado com outras meta-heurísticas encontradas na literatura tais como AG e SA. Para o PCV as instancias utilizadas foram Oliver 30 (Oliver et al, 1987), Eil 50 e Eil 75 (Eilon et al, 1971) e Kro 100 da base de dados TSPLIB. Os resultados obtidos mostraram que ACO é mais eficiente. O ACO é sem dúvida um clássico algoritmo guloso, da mesma forma que o GRASP. Assim sugere-se também que a investigação de um algoritmo baseado no GRASP pode ser uma fonte de informação importante para se projetar um algoritmo baseado em ACO.

3 MATERIAL E MÉTODOS

Nesta seção são abordadas as técnicas aplicadas para a obtenção dos indicadores de desempenho utilizados neste estudo. Será feita uma introdução, seguida de algumas definições e conceitos relacionados à teoria de filas com distribuições exponenciais, bem como o modelo hipercubo de filas desenvolvido por Larson (1974), relacionando com o caso em que o processo das chegadas dos chamados é uma distribuição de Poisson e há vários servidores com tempos de serviços de atendimento exponencialmente distribuídos. Também são estabelecidas algumas definições e conceitos sobre a meta-heurística *Ant Colony Optimization* (ACO), originalmente utilizada no problema do caixeiro viajante (PCV). Finalmente, é proposto um algoritmo baseado na meta-heurística ACO, com o objetivo de gerar novos átomos geográficos e, com isso, comparar as medidas de desempenho obtidas por meio do modelo com as duas divisões: a primeira, estabelecida segundo os Batalhões da Polícia Militar (BPM) de Curitiba, e a divisão proposta com a utilização da heurística proposta neste trabalho.

3.1 TEORIA DAS FILAS

Teoria de filas é um ramo da probabilidade que trata de filas por análises matemáticas precisas. De acordo com Moreira (2011), “teoria das filas é um corpo de conhecimentos matemáticos, aplicado ao fenômeno das filas.” É uma das técnicas da pesquisa operacional que trata de assuntos relacionados aos problemas de congestionamento de sistemas que ofereçam serviços cuja demanda cresce de maneira aleatória e que, de alguma forma, podem gerar filas devido a restrições no atendimento. Uma possível solução para esse problema é a de encontrar um equilíbrio entre as necessidades do cliente e a capacidade do provedor do serviço, para buscar a satisfação com o menor custo possível (MORAES, 2011).

A teoria de filas pode ser aplicada nos mais diversos serviços básicos de atendimento que a população utiliza, entre eles pode-se citar: atendimento médico emergencial, policiamento, transportes coletivos, coleta de lixo, entre outros. À medida que a população aumenta, o que nem sempre ocorre de maneira linear, faz-se necessário melhorar esse tipo de atendimento de serviço básico. A teoria de filas apresenta-se como uma poderosa ferramenta matemática capaz de prover esta melhoria.

Os diferentes tipos de serviços de atendimentos básicos oferecidos à população, conforme descrito no item 2.3, dividem-se em três categorias: a primeira classifica-se como serviços de rotina; a segunda, como serviços semiemergenciais e a terceira, como serviços emergenciais.

Ao analisar os três tipos de serviços prestados à sociedade, percebe-se que nos dois primeiros tipos de atendimento, a existência de pequenos atrasos no tempo de atendimento não acarretará grandes problemas à população. No entanto, no terceiro tipo de atendimento, atrasos de qualquer natureza podem ser muito prejudiciais. Segundo Marinho (2006), o tempo perdido em filas podem exercer impactos significativos sobre o bem estar, as probabilidades de cura, a natureza e extensão das sequelas nos pacientes, nos familiares envolvidos e na sociedade como um todo.

3.1.1 Histórico

A abordagem matemática da teoria de filas iniciou-se em 1908, na Dinamarca, cidade de Copenhague. Quem iniciou o estudo matemático do modelo foi Agner Krup Erlang, em 1909, que trabalhava numa empresa de telefonia e tentava solucionar o problema de atendimento das centrais telefônicas. Porém, o estudo da teoria de filas só foi aplicado em outros casos relacionados ao congestionamento, após a Segunda Guerra Mundial. As pesquisas, nesse sentido, começaram a se aprofundar na década de 1930, quando Andrey Kolmogorov, na Rússia, resolveu estudar problemas com entradas de probabilidade de Poisson e saídas arbitrárias com um ou vários atendentes. Essa técnica pode ser encontrada em diversas áreas da literatura, tais como Pesquisa Operacional, Probabilidade e também na Engenharia Industrial. Destacam-se sua aplicabilidade no fluxo de tráfego, considerando carros, aviões, pessoas, entre outros; no escalonamento, como por exemplo, pacientes em hospitais e programas em computadores; e também na prestação de serviços, tais como, o atendimento aos usuários de banco, o atendimento do corpo de bombeiros, e serviço de atendimento médico de urgência e emergência entre outros.

3.1.2 Definições de Teoria de Filas

Muitas definições foram elaboradas desde que a Teoria de Filas surgiu como uma nova perspectiva para problemas de congestionamento, com objetivo de otimizar o desempenho do sistema e reduzir seus custos operacionais. Há diversos tipos de filas e componentes. Segundo Prado (2006), as filas contêm os seguintes elementos: população, entidade, servidor e serviço.

Para melhor entendimento de um sistema de filas e seus componentes, pode-se visualizar a (FIGURA 3.1), um dos exemplos de fila, com uma população aleatória à espera de um serviço em uma única fila e com apenas um servidor na unidade de atendimento.

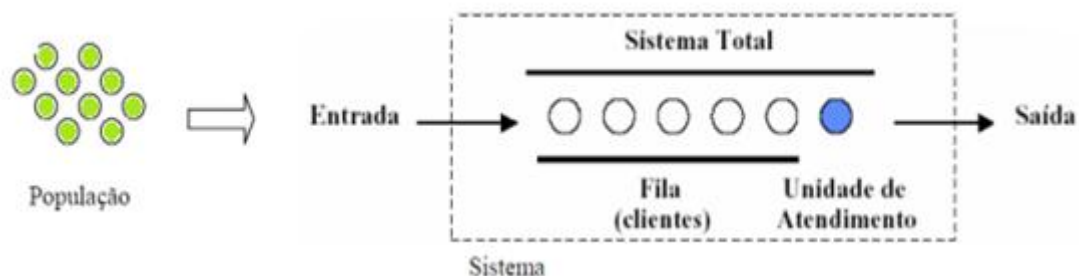


FIGURA 3.1 - EXEMPLO DE FILA COM SEUS COMPONENTES
FONTE: Marins (2012)

São apresentadas a seguir, algumas definições básicas relacionadas à Teoria de Filas, baseado em Prado (2006).

- **Tamanho da população:** Diz respeito ao número de clientes que podem chegar a um sistema, e pode ser finito ou infinito. Se este número for inferior a 30, diz-se que este grupo é pequeno, se o número de clientes for maior que 30, será considerado infinito, e neste caso a chegada de um novo cliente não afetará significativamente a probabilidade de chegada de outro cliente.
- **Clientes:** São unidades da população em estudo que chegam para o atendimento, podendo ser pessoas, peças, automóveis, máquinas em geral. Segundo Fogliatti e Costa (2006), a chegada de clientes no sistema é classificada como individual ou em grupos. Para a chegada individual, tem-se como exemplo a chegada de um navio em um porto de carregamento. Um exemplo de chegada em grupo são pessoas de uma excursão chegando a um ponto turístico.

- Fila: Número de pessoas esperando o atendimento.
- Unidade de atendimento: Processo ou sistema que realiza o atendimento. Segundo Hillier e Lieberman (2001), o atendimento de um sistema de fila se dá por meio de um ou mais atendentes em paralelo. Com isso, a unidade de atendimento pode ser única ou múltipla. E ainda o número de canais de atendimento pode ser finito ou infinito (Fogliatti e Costa, 2006).
- Taxa de chegada dos clientes (λ): É o valor esperado do número de chegadas de clientes por unidade de tempo. Aqui são utilizadas distribuições de frequências para representar as chegadas dos clientes, uma vez que a taxa de chegada ocorre de forma aleatória.
- Taxa de atendimento dos clientes (μ): Número de clientes por unidade de tempo. É conhecido como a razão de atendimento. O valor dessa taxa de atendimento é considerado como se o atendente estivesse ocupado todo o seu tempo, ou seja, sem tempo ocioso. Também aqui, devido à irregularidade dos atendimentos, necessita-se descobrir qual distribuição de frequência melhor se enquadra. O pressuposto mais comum é a distribuição de Poisson. Dessa forma, faz-se necessário que os eventos de chegadas sejam independentes. Os resultados obtidos em todos os casos são valores médios e supõe-se que as taxas se mantêm constantes ao longo do período.
- Disciplina da fila: É o método para decidir qual o cliente que será atendido primeiro.
- Número médio de clientes na fila não vazia (NF): Número médio de clientes aguardando o atendimento é a característica mais relevante para a escolha da fila.
- Número médio de clientes no sistema (NS): Número de clientes aguardando na fila mais o(s) que está (ão) sendo atendido(s).
- Tempo médio de cliente na fila (TF): Tempo médio de espera na fila.
- Tempo médio do cliente no sistema (TS): Tempo médio na fila mais o tempo de atendimento. A partir do número médio de clientes no sistema ou na fila, é possível calcular o tempo médio de permanência do cliente no sistema (TS) e na fila (TF).
- Fator de utilização do servidor (ρ): Representa a fração média do tempo em que o servidor está ocupado. É a base de cálculo da probabilidade de haver

um número k de clientes no sistema. Neste caso definirá o tamanho da fila e o tempo médio de permanência dos clientes no sistema ou na fila.

Ao se considerar um problema de filas, em que a taxa de atendimento dos clientes seja maior que a taxa de chegada dos clientes ($\mu \geq \lambda$), pode-se concluir erroneamente que não existe fila no sistema. Porém, devido à aleatoriedade dos processos de chegada e de atendimento, não se pode afirmar a inexistência de filas no sistema.

3.1.3 Funcionamento do sistema

Um sistema é considerado congestionado quando suas filas ultrapassam um valor pré-estabelecido ou um valor que é considerado o normal para o sistema. Neste momento, o sistema poderá apresentar baixa produtividade e tempo de espera excessivo ao usuário. São diversos os fatores que podem influenciar para um congestionamento no sistema, entre eles pode-se citar:

- Forma com que os funcionários atendem os clientes;
- Forma de chegada dos clientes;
- Disciplina da fila;
- Estrutura do sistema.

3.1.3.1 Forma com que os funcionários atendem os clientes

A maneira como os funcionários atendem os clientes pode influenciar no tempo de atendimento. Para analisar um sistema de filas é preciso coletar alguns dados, entre eles, encontrar estatisticamente o número de clientes atendidos em uma unidade de tempo, bem como o tempo gasto em cada atendimento. Esse procedimento ajudará a encontrar a distribuição estatística de probabilidade do número de atendimentos e o tempo gasto em cada atendimento.

3.1.3.2 Forma de chegada dos clientes

A forma de chegada dos clientes ocorre de maneira totalmente aleatória. Dessa forma, faz-se necessário um levantamento estatístico para verificar se é

possível encontrar uma distribuição de frequência que se enquadre na forma de chegada dos clientes. Essa distribuição auxiliará na tomada de decisão por parte do gerente do sistema.

Para efetuar a coleta desses dados é necessário que o processo de chegada esteja no estado estacionário, indicando que o processo poderá ser sempre representado por este tipo de levantamento de dados. Um exemplo de estado estacionário são os usuários de uma agência de correios, que em dias normais, utilizam-na em um processo estacionário, porém quando existe a possibilidade de uma greve prolongada nos correios, o sistema poderia ser classificado como não estacionário, pois haveria uma corrida dos usuários ao Correio. Situações desse tipo podem influenciar nas características das filas, e com isso alterar a distribuição de probabilidade.

3.1.3.3 A disciplina da fila

Disciplina da fila é um conjunto de limitações que impõem a ordem da organização da fila na chegada dos clientes, ou seja, quem será atendido em primeiro lugar. Por vezes, poderá ser o primeiro a chegar à fila, ou o último a chegar à fila, ou ainda o que exige um atendimento especial, com características especiais, entre outros fatores. Alguns tipos de disciplina de filas são:

- FIFO (*First-In-First-Out*) ou FCFS (*First Come, First Served*): o primeiro cliente a chegar na fila será atendido.
- LIFO (*Last-In-First-Out*) ou LCFS (*Last Come, First Served*): o último cliente a chegar na fila será atendido.
- SIRO (*Service_In-Random_Order*): o atendimento dos clientes faz-se por ordem aleatória.
- SPT (*Shortest-Processing_Time-First*): será atendido em primeiro lugar o cliente com menor tempo de atendimento.
- PR (*Priority Rules*): o atendimento deve obedecer à prioridade pré-estabelecida, de acordo com algumas regras do sistema.

A notação de um sistema de filas, normalmente é descrita segundo a notação de Kendall (1953). Possui seis letras como no exemplo a seguir:

$(A/b/ C/ k/ M/ z)$

Em que:

$A \rightarrow$ especifica a distribuição dos intervalos entre as chegadas;

$b \rightarrow$ especifica a distribuição do tempo de serviço;

$C \rightarrow$ o número de canais de atendimento ou número de servidores

$k \rightarrow$ especifica o número máximo (capacidade máxima) de usuários no sistema.

$M \rightarrow$ fornece o tamanho da população do sistema

$z \rightarrow$ descreve a disciplina da fila.

Temos ainda, de acordo com as características “A” e “b”, as abreviações padrões dos tipos de distribuições:

M - intervalos de tempo entre chegadas - são independentes, identicamente distribuídos e variáveis aleatórias, seguindo o modelo Markoviano (distribuição exponencial negativa ou distribuição de Poisson);

D - intervalos de tempo entre chegadas - são independentes, identicamente distribuídos e determinísticos (distribuição determinística);

Em - intervalos de tempo entre chegadas - são independentes, identicamente distribuídos e variáveis aleatórias tendo distribuição de Erlang de ordem “m”;

Hm - hiper-exponencial de estágio “m”;

G - intervalos de tempo entre chegadas - são independentes, identicamente distribuídos e com distribuição genérica.

Há alguns casos particulares de notação, como por exemplo, a notação $A/B/c$. Ela supõe que não há limites para o tamanho da fila, a população é infinita e a

disciplina da fila é FIFO. Se houver restrições para a capacidade, a notação passa a ser A/B/c/K.

Os modelos de distribuição de Poisson e Markovianos são os mais utilizados, pois é possível encontrarem-se as principais características da fila sem análises elaboradas com alta complexidade matemática.

3.1.3.4 A estrutura do sistema

Há vários tipos de estrutura do sistema que necessitam ser estudadas caso a caso. A seguir estão relacionados três exemplos de configurações, conforme mostra a (FIGURA 3.2).

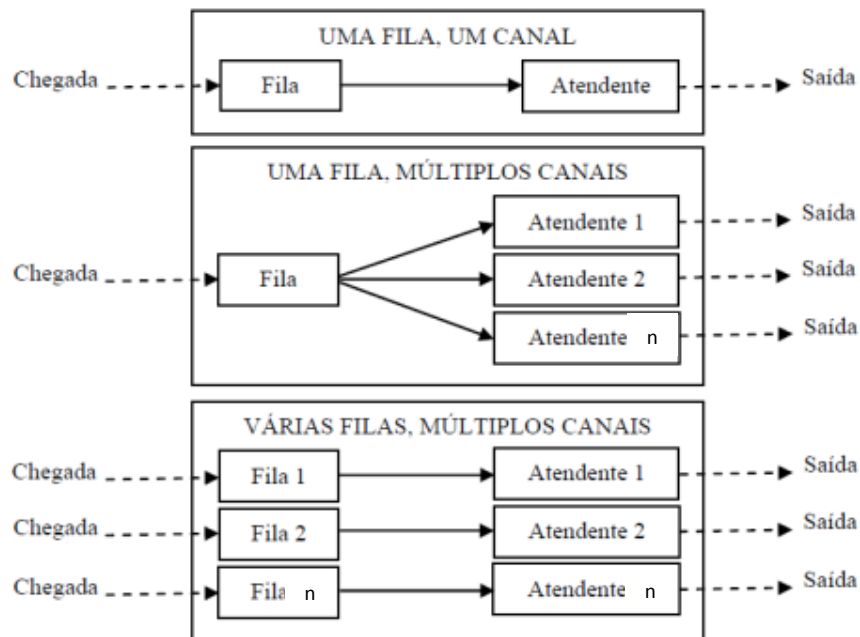


FIGURA 3.2 ESTRUTURAS DOS SISTEMAS DE FILAS

FONTE: Camelo *et al* (2010)

Tem-se, na figura acima, inicialmente, uma única fila e um único canal de atendimento. Em seguida existe uma fila com múltiplos canais de atendimento. Finalmente, tem-se um exemplo de sistema mais complexo, com várias filas e múltiplos canais. Esses são alguns dos exemplos de fila. A literatura sobre esse assunto é bem ampla e diversificada.

3.1.4 Modelo $M/M/1$ e $M/M/m$

O modelo $M/M/1$, conforme a notação de Kendall (1953), considerado o mais simples na literatura da pesquisa operacional, representa um sistema com apenas 1 atendente e uma fila com população infinita. Os intervalos entre as chegadas e os tempos de atendimento representam distribuições Markovianas (M), quer dizer, o número de chegadas por intervalo de tempo possui distribuição de Poisson e a distribuição dos tempos de atendimento é exponencial. As chegadas e os tempos de serviço possuem médias $E(X) = \lambda$ e $E(S) = 1/\mu$, respectivamente. Considerando a taxa de chegada de clientes por intervalo de tempo igual a λ e a taxa média de atendimento por intervalo de tempo igual a μ , determina-se o índice de congestionamento para o sistema igual a ρ , onde $\rho = \lambda/\mu$. Caso o fator de utilização do servidor seja maior que um, significa que o sistema está congestionado. Nesse modelo, não existe margem prática para dimensionamento do sistema. Desta forma, o controle do sistema é realizado apenas aumentando ou diminuindo a quantidade dos servidores.

No caso de vários servidores, modelo $M/M/m$, o m representa o número de canais de atendimento, as mesmas hipóteses do caso anterior são válidas, e considera-se que todos os servidores possuirão a mesma taxa de serviço. Os usuários do sistema esperam em uma única fila, com capacidade infinita, e são atendidos pela disciplina FCFS (*First Come First Served*). O fator de utilização $\rho = \lambda/(\mu.m)$ corresponde à taxa média de utilização do sistema (considerando $\rho < 1$, caso contrário o sistema nunca entraria em equilíbrio). Nesse modelo não será considerada a variação na taxa de chegada de usuários (λ), contudo, poderão existir mudanças nas taxas de serviços de atendimento (μ), ou seja,

$$\lambda_n = \lambda \quad \text{para } n = 0, 1, 2, \dots \quad (3.1)$$

$$\mu_n = n\mu, \quad \text{para } n = 1, 2, \dots, m-1 \quad (3.2)$$

$$\mu_n = m\mu, \quad \text{para } n = m, m+1 \quad (3.3)$$

3.1.5 Transição de estado

Transição de estado é a passagem de um estado para outro, por exemplo, um servidor que estava ocupado e termina seu atendimento passando para o estado livre, ou vice-versa. A transição depende apenas do estado atual do sistema, e não da história do processo das transições.

A fim de ilustrar a transição de estado do modelo $M/M/m$ de teoria de fila, segundo Souza (2010), serão apresentados dois exemplos ilustrativos. No primeiro exemplo, tem-se uma fila que o número de usuários (n) é menor do que o número de atendentes (m). No segundo exemplo, o número de usuários é maior ou igual a do número de atendentes.

A fim de facilitar a análise dos exemplos ilustrativos das transições de estado mostrado a seguir, não será considerada a chegada de mais de um cliente por intervalo de tempo, mas apenas a chegada de um cliente por vez.

No primeiro exemplo, tem-se uma situação conforme mostrado na (FIGURA 3.3), em que o número de usuários é menor do que o número de atendentes do sistema ($n < m$), pois cada cliente que chega ao sistema é prontamente atendido.

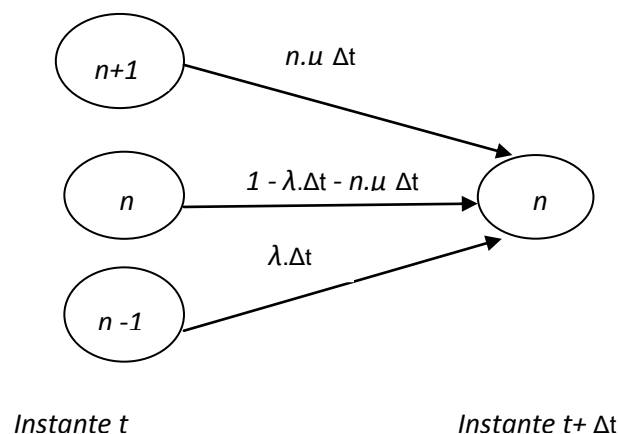


FIGURA 3.3 - TRANSIÇÕES DE ESTADO DO MODELO $M/M/m$, QUANDO $n < m$
 FONTE: Souza (2010)

De acordo com a FIGURA 3.3, têm-se as seguintes transições de estado do modelo $M/M/m$, lembrando que todo usuário é atendido imediatamente. O estado $n+1$ (no instante t) muda para o estado n com o término de um atendimento (Δt), com probabilidade $n \cdot \mu \cdot \Delta t$. Da mesma forma, estando no estado $n-1$ (no instante t) muda para o estado n somente com a chegada de uma nova chamada (Δt), com probabilidade $\lambda \cdot \Delta t$ (menos para o estado $n = 0$). Por último, se o sistema encontra-

se no estado n no instante t e não acontece nenhuma chamada ou término de atendimento durante o tempo Δt , o sistema permanecerá no estado n com probabilidade igual a $1 - \lambda \Delta t - n \mu \Delta t$.

Na segunda situação, conforme podemos ver na (FIGURA 3.4), $n \geq m$, todos os servidores poderão estar ocupados. Neste caso, a chegada de uma nova chamada ao sistema terá que aguardar em uma fila de espera.

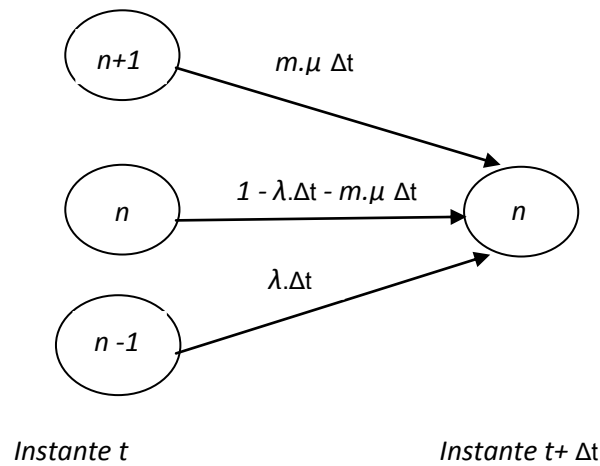


FIGURA 3.4 - TRANSIÇÕES DE ESTADO DO MODELO M/M/m QUANDO $n \geq m$
 FONTE: Souza (2010)

Nesse tipo de transição, o estado $n + 1$ (tempo t), somente passará para o estado n com o término de um atendimento durante o intervalo de tempo Δt , com a probabilidade de $m \cdot \mu \cdot \Delta t$. Agora, se o sistema encontra-se no estado n num instante t e não acontece nenhuma chamada ou término de atendimento durante o intervalo de tempo Δt , o sistema permanecerá no estado n com probabilidade igual a $1 - \lambda \cdot \Delta t - m \cdot \mu \cdot \Delta t$. Por fim, a mudança de um estado $n - 1$ para o estado n só ocorrerá com a chegada de uma nova chamada ao sistema, com probabilidade de $\lambda \cdot \Delta t$.

Portanto, a probabilidade do sistema encontrar-se no estado n no instante $t + \Delta t$ é dada por:

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = \mu P_1(t) - \lambda P_0(t) + \frac{o(t)}{\Delta t} \quad n = 0 \quad (3.4)$$

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -(\lambda + n\mu)P_n(t) + (n+1)\mu.P_{n+1}(t) + \lambda.P_{n-1}(t) + \frac{o(t)}{\Delta t} \quad n = 1, \dots, m-1 \quad (3.5)$$

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -(\lambda + m\mu)P_n(t) + m.\mu.P_{n+1}(t) + \lambda.P_{n-1}(t) + \frac{o(t)}{\Delta t} \quad n = m, m+1, \dots \quad (3.6)$$

Logo, com Δt tendendo a zero, tem-se:

$$\frac{dP_0(t)}{dt} = \mu P_1(t) - \lambda P_0(t) \quad n = 0 \quad (3.7)$$

$$\frac{dP_n(t)}{dt} = -(\lambda + \mu)P_n(t) - (n+1)\mu.P_{n+1}(t) + \lambda.P_{n-1}(t) \quad n = 1, \dots, m-1 \quad (3.8)$$

$$\frac{dP_n(t)}{dt} = -(\lambda + m\mu)P_n(t) + m.\mu.P_{n+1}(t) + \lambda.P_{n-1}(t) \quad n = m, m+1, \dots \quad (3.9)$$

As equações (3.7), (3.8) e (3.9) analisam o comportamento transiente do sistema de filas ao longo do tempo.

Logo, P_n é a probabilidade do sistema encontrar-se em equilíbrio no estado n independentemente do tempo t ($\lim_{t \rightarrow \infty} P_n(t) = P_n$), desconsiderando o período transiente que ocorre durante o início da operação do sistema. Fazendo: $\frac{dP_n(t)}{dt} = 0$

têm-se as probabilidades e com isso o sistema que, inicialmente apresentava infinitas equações diferenciais, se reduzirá a um sistema de equações lineares nas variáveis P_n :

$$\lambda P_0 = \mu P_1 \quad n = 0 \quad (3.10)$$

$$(\lambda + \mu)P_1 = \lambda P_0 + 2\mu P_2 \quad n = 1 \quad (3.11)$$

$$\begin{array}{ccc} \vdots & & \vdots \\ (\lambda + n\mu)P_n = \lambda P_{n-1} + (n+1)\mu P_{n+1} & n = 3, \dots, m & (3.12) \end{array}$$

$$(\lambda + m\mu)P_m = \lambda P_{m-1} + m\mu P_{m+1} \quad n = m, m+1, \dots \quad (3.13)$$

Pode-se verificar nas equações (3.10), (3.11), (3.12) e (3.13) que em média as taxas de saída do sistema são iguais às taxas de chegada ao sistema. Dessa forma, as equações podem ser reescritas da seguinte maneira:

$$P_1 = \frac{\lambda}{\mu} P_0 \quad (3.14)$$

$$P_2 = \frac{\lambda P_1 + \mu P_1 - \lambda \frac{\mu}{\lambda} P_1}{2\mu} = \frac{\lambda}{2\mu} P_1 \quad (3.15)$$

$$\begin{array}{ccc} \vdots & & \\ P_n = \frac{\lambda P_{n-1} + \mu P_{n-1} - \lambda \frac{\mu}{\lambda} P_{n-1}}{n\mu} = \frac{\lambda}{n\mu} P_{n-1} & & (3.16) \end{array}$$

Cabe ressaltar que o recurso utilizado a fim de obter as equações (3.10), (3.11), (3.12) e (3.13) são válidas apenas para filas em que os estados se comunicam de dois em dois, em uma fila simples, conforme a (FIGURA 3.5).

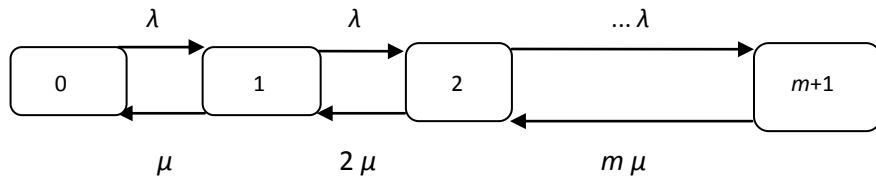


FIGURA 3.5 TRANSIÇÃO DE ESTADOS DO MODELO $M/M/m$
FONTE: AUTOR

Para modelos os $M/M/m/C$, existe uma restrição quanto ao número de usuários, utiliza-se raciocínio análogo. Porém, a limitação do número de usuários implica na flexibilidade quanto ao índice de congestionamento, podendo este ser

maior que 1. Cabe ainda ressaltar que, para sistemas estáveis, se ρ tender a 1, a fila tende a aumentar infinitamente.

3.2 RELAÇÃO ENTRE A DISTRIBUIÇÃO DE POISSON E A DISTRIBUIÇÃO EXPONENCIAL

3.2.1 Distribuição de Poisson

A distribuição de Poisson é uma distribuição discreta de probabilidade aplicável a diversos tipos de ocorrências relacionadas a eventos com intervalos de tempos previamente determinados, conhecidos como taxas de chegada de ocorrências (λ). Pode-se citar como exemplos: clientes chegando ao caixa de um supermercado; número de automóveis que chegam a um posto de combustível; número de aviões pousando no aeroporto em um determinado período; chegada de chamados de emergências aos SAEs, entre outras.

A distribuição de Poisson descreve basicamente o número de eventos ($N(t)$) que ocorre num determinado intervalo de tempo.

O processo de Poisson começa no instante zero com $N(0) = 0$, que representa um sistema vazio, como é definido por Feller (1968). Definem-se dois importantes postulados relacionados a esse processo.

- As transições a partir de um estado n são possíveis apenas para um estado posterior “ $n+1$ ”;
- Qualquer que seja o estado n num determinado instante t , a probabilidade de ocorrência de um evento dentro de um pequeno intervalo t e $t + \Delta t$ é igual a $\lambda \cdot \Delta t$, no qual λ é a taxa de chegada. A probabilidade de ocorrência de mais de um evento nesse intervalo de tempo é $o(\Delta t)$, tal que:

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0 \quad (3.17)$$

Considerando um exemplo em que o sistema encontre-se num estado n , e que não ocorra nenhum evento no intervalo de tempo $(t \text{ e } t + \Delta t)$, pode-se dizer que a probabilidade será igual a

$$P_n(t) \cdot [1 - \lambda \cdot \Delta t] + o(\Delta t) \quad (3.18)$$

Caso o sistema encontre-se num estado anterior $n-1$ no instante t e ocorra um evento nesse mesmo intervalo de tempo $(t \text{ e } t + \Delta t)$, o sistema automaticamente mudará para o próximo estado n no tempo $t + \Delta t$ com probabilidade de

$$P_{n-1}(t) \cdot \lambda \cdot \Delta t \quad (3.19)$$

Dessa forma, a probabilidade do sistema encontrar-se no estado n no instante $t + \Delta t$, será igual a

$$P_n(t + \Delta t) = P_n(t) \cdot [1 - \lambda \cdot \Delta t] + P_{n-1}(t) \lambda \cdot \Delta t + o(\Delta t) \quad (3.20)$$

Logo, para o caso particular de $n = 0$, a probabilidade de o sistema encontrar-se no estado n no intervalo de tempo $(t + \Delta t)$, é

$$P_0(t + \Delta t) = P_0(t) \cdot [1 - \lambda \cdot \Delta t] + o(\Delta t) \quad (3.21)$$

Reescrevendo as equações 3.20 e 3.21, tem-se:

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = \lambda \cdot P_n + \lambda \cdot P_{n-1}(t) + \frac{o(\Delta t)}{\Delta t} \quad (3.22)$$

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = \lambda \cdot P_0 + \frac{o(\Delta t)}{\Delta t} \quad (3.23)$$

No limite quando $\Delta t \rightarrow 0$ tem-se:

$$\left\{ \begin{array}{l} \frac{dP_0(t)}{dt} = -\lambda.P_0(t), \quad \text{para } n = 0 \\ \frac{dP_n(t)}{dt} = -\lambda.P_n(t) + \lambda.P_{n-1}(t), \quad \text{para } n \geq 1 \end{array} \right. \quad (3.24)$$

$$\left\{ \begin{array}{l} \frac{dP_n(t)}{dt} = -\lambda.P_n(t) + \lambda.P_{n-1}(t), \quad \text{para } n \geq 1 \end{array} \right. \quad (3.25)$$

Dessa forma, segundo as equações (3.24) e (3.25), para $n = 0$, tem-se que:

$$P_0(t) = e^{-\lambda t} \quad (3.26)$$

Para $n = 1$, tem-se $P_1(t) = (\lambda t)e^{-\lambda t}$, logo, por indução, encontra-se:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \text{ para } n = 0, 1, 2, \dots \quad (3.27)$$

A partir dos postulados citados anteriormente, tem-se na equação (3.27), a distribuição de Poisson. Logo, constata-se a veracidade da probabilidade de não ocorrer nenhuma chamada até o instante t , dada por $P_0(t) = e^{-\lambda t}$, que é exatamente a probabilidade de que a primeira chegada ocorra depois do instante t em uma distribuição exponencial ($P(T > t)$).

$$P_0(t) = P(T > t) = e^{-\lambda t} \quad (3.28)$$

A partir da equação 3.28, tem-se uma importante relação entre a distribuição de Poisson e a distribuição exponencial, conclui-se que se o número de vezes que um evento ocorre ao longo do tempo possui uma distribuição de Poisson, os intervalos de tempo entre as ocorrências consecutivas dos eventos seguem uma distribuição exponencial. Essa relação é útil para a construção das equações de equilíbrio das filas Markovianas. Nesse estudo, em especial, essa relação é utilizada para a comprovação da segunda hipótese do modelo hipercubo de filas, em que os chamados de atendimento chegam a partir de um processo de Poisson.

3.2.2 Distribuição Exponencial

A distribuição exponencial é uma distribuição contínua de probabilidade e é representada pelo parâmetro da distribuição λ , muito utilizada para modelar o tempo

entre ocorrências de eventos num processo de Poisson. Pode-se expressar sua função de densidade de duas maneiras conforme mostra as equações (3.29) e (3.30):

$$\begin{cases} f(x; \lambda) = \lambda e^{-\lambda x}, & x \geq 0 \\ f(x; \lambda) = 0, & x < 0 \end{cases} \quad (3.29)$$

$$\begin{cases} f(x; \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, & x \geq 0, \lambda > 0 \\ f(x; \lambda) = 0, & x < 0 \end{cases} \quad (3.30)$$

De acordo com as equações (3.29) e (3.30), a probabilidade da variável x assumir valor negativo é nula. E a probabilidade da variável x assumir valores positivos é $\lambda \cdot e^{-\lambda x}$.

A média e a variância são dadas por:

$$E(X) = \frac{1}{\lambda} \quad V(X) = \frac{1}{\lambda^2}$$

As distribuições exponenciais apresentam diversas ramificações e aplicabilidade.

Segundo Souza (2010), em uma distribuição exponencial a probabilidade de uma variável aleatória contínua T ser maior que $t+s$, com T maior que s , não depende de s como definida a seguir:

$$P(T > s+t / T > s) = \frac{P(T > s+t \cap T > s)}{P(T > s)} \quad (3.31)$$

$$\text{Como:} \quad P(T > s+t \cap T > s) = P(T > s+t) \quad (3.32)$$

$$\text{Então:} \quad P(T > s+t / T > s) = \frac{P(T > s+t)}{P(T > s)} \quad (3.33)$$

$$P(T > s + t / T > s) = \frac{e^{-(s+t)\lambda}}{e^{-s\lambda}} = e^{-t\lambda} = P(T > t) \quad (3.34)$$

Esta é uma propriedade da distribuição exponencial conhecida como “falta de memória”, utilizada para determinar, por exemplo, a probabilidade de um servidor terminar seu atendimento antes dos demais servidores, caso todos estejam ocupados e considerando que os chamados respeitem o processo de Poisson, conforme apresentado na equação (3.34). É uma das propriedades das filas Markovianas que também será utilizada no presente trabalho.

Portanto, a distribuição de Poisson e a distribuição exponencial estão relacionadas. A distribuição de Poisson estima a quantidade de eventos num determinado intervalo de tempo. É uma distribuição de dados discreta, enquanto que a distribuição exponencial analisa o intervalo de tempo para a ocorrência do evento. Logo, se a distribuição entre os tempos das chamadas emergenciais for exponencial, pode-se concluir que o número de chamadas emergenciais se enquadra numa distribuição de Poisson.

3.3 MODELO HIPERCUBO DE FILAS

O modelo hipercubo de filas tem por essência a Teoria de Filas e baseia-se na teoria de Markov em tempo contínuo. O modelo não determina uma configuração ótima para o sistema, pois não é um modelo de otimização, no entanto, pelo fato de ser uma ferramenta que permite calcular diversos indicadores de desempenho, fornece uma completa avaliação de desempenho de cada configuração sugerida pelo avaliador que podem auxiliar nas decisões operacionais do sistema. Foi desenvolvido por Larson (1974), e aperfeiçoado por diversos autores. Foi proposto para descrever sistemas em que servidores deslocam-se para oferecer algum tipo de serviço (*server-to-customer service*). Inicialmente, foi aplicado para resolver um problema de patrulhamento policial, nos Estados Unidos. É um modelo estocástico descritivo e considera toda a complexidade geográfica e temporal de um sistema. O modelo hipercubo de filas pode analisar sistemas com apenas um ou diversos atendentes, sendo que estes podem estar espacialmente distribuídos ao longo de uma região. O modelo é utilizado tanto em sistemas que admitem filas (caudas) quanto nos que não admitem filas de espera.

O modelo hipercubo de filas baseia-se fundamentalmente na partição de uma região em átomos geográficos independentes, onde cada átomo se comporta como uma fonte individual e pontual. Na prática, a região a ser estudada deve ter condições de ser particionada em regiões menores, ou seja, um conjunto finito de áreas geradoras de demanda. O atendimento poderá ser realizado por diversos servidores (veículos) distribuídos ao longo de uma região. A localização dos servidores deverá ser conhecida, ao menos probabilisticamente (CHAIKEN, 1975).

Para esse modelo não basta apenas especificar a quantidade de servidores ocupados, mas conhecer a disponibilidade de cada um. Para isso, utiliza-se uma variável binária associada a cada servidor, com os valores 0 ou 1 representando os estados, livre ou ocupado, respectivamente, do servidor, num determinado instante. O atendimento é realizado preferencialmente pelo servidor principal. Caso esteja ocupado, o atendimento pode ser realizado pelo servidor que estiver disponível (*backup*). Caso não haja servidor disponível, o atendimento é perdido ou entra em fila de espera, dependendo do sistema.

O nome hipercubo é derivado do espaço de estados que descreve o *status* dos servidores, uma vez que eles podem estar livres (0) ou ocupados (1). Por exemplo, para um sistema com três servidores, o estado representado por (001) corresponde à situação onde o primeiro servidor está ocupado, o segundo e o terceiro servidores estão livres, os servidores estão representados da direita para a esquerda. Desta maneira, um sistema com três servidores tem seu espaço de estados em R^3 , e é dado pelo conjunto de vértices do cubo de arestas iguais a 1 conforme apresentado na (FIGURA 3.6). A medida que cresce o número de servidores, a representação se dá no espaço N-dimensional, pelo conjunto de vértices de um hipercubo, daí o nome hipercubo de filas.

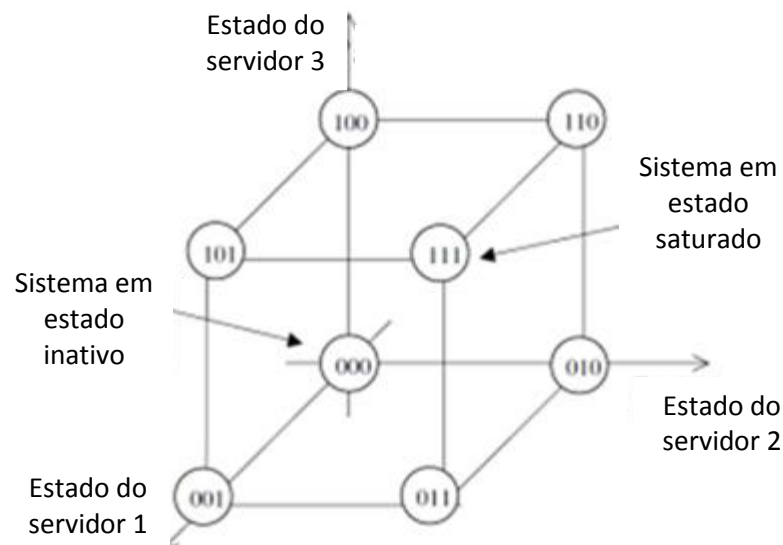


FIGURA 3.6 - REPRESENTAÇÃO DO ESPAÇO DE ESTADOS PARA UM SISTEMA DE ATENDIMENTO ESPACIALMENTE DISTRIBUÍDO, COM TRÊS SERVIDORES

FONTE: Costa (2003)

Considerando um exemplo hipotético com três átomos ($N_A = 3$) e três servidores ($N = 3$), como mostra a figura anterior, têm-se ao todo oito possíveis estados, sem considerar fila de espera. O estado (1,1,0), por exemplo, corresponde ao servidor número 1 estar livre e os servidores 2 e 3 estarem ocupados.

3.3.1 Hipóteses do Modelo Hipercubo de Filas

Para que o modelo hipercubo de filas possa ser aplicado ao Sistema de Atendimento Emergencial (SAE) móvel, as seguintes hipóteses, segundo Larson e Odoni (1981), devem ser satisfeitas:

1. Divisão da área de atendimento em N_A átomos geográficos independentes:

A área de atendimento deverá ser dividida em N_A áreas menores independentes, áreas que se complementem geograficamente. A partição da região em átomos geográficos independentes, em que cada átomo se comporta como uma fonte individual e independente é uma hipótese fundamental para que o modelo possa ser aplicado a um sistema. Há diversas maneiras de fazer essa divisão, como por exemplo, setores policiais, bairros, divisão política, entre outras.

2. As solicitações por serviço geradas em cada átomo j ($j = 1, \dots, N_A$), devem chegar independentemente de acordo com uma distribuição de Poisson.

É necessário verificar estatisticamente se o processo de chegada dos chamados em cada átomo j ocorre de maneira independente, com taxas λ_j , possíveis de serem medidas ou estimadas. Apesar dessa hipótese inicialmente parecer restritiva, ela é normalmente satisfeita nos serviços que envolvem aleatoriedade espacial e geográfica. É facilmente comprovada por meio dos testes de estatísticos, Qui-quadrado e Kolmogorov-Smirnov.

3. Os tempos de viagem dos veículos do átomo i para o átomo j ($i, j = 1, \dots, N_A$) devem ser conhecidos ou estimados.

Aqui, consideram-se os tempos médios de viagem entre os átomos i, j ($i, j = 1, 2, 3, \dots, N_A$). Essa hipótese também é facilmente satisfeita, os sistemas normalmente dispõem desses tempos. Caso contrário, esses tempos podem ser estimados por conceitos de probabilidade geométrica.

4. O sistema deverá operar com servidores espacialmente distribuídos, que poderão se deslocar para atender a qualquer um dos átomos.

Essa hipótese trata da possibilidade de todos os servidores atenderem a todos os átomos. Contudo, é uma hipótese que pode ser flexibilizada, segundo Mendonça (1999), a fim de atender particularidades de políticas de despachos.

5. A localização dos servidores deve ser conhecida, ao menos probabilisticamente.

No caso de servidores com bases fixas, sua localização torna-se simples, a localização é a sua base. Contudo, para servidores que se movem dentro de uma determinada área de atuação, a sua localização deverá ser estimada probabilisticamente. A representação das localizações dos servidores é feita de forma matricial $L = [l_{nj}]$, na qual os elementos da matriz representam a probabilidade de um servidor estar localizado em determinado átomo no momento em que se encontrar disponível.

6. Despacho ou alocação de um servidor.

Apenas um servidor deverá ser despachado para atender um chamado. O modelo hipercubo de filas não representa adequadamente sistemas onde são despachados mais de um servidor para atender o mesmo chamado.

7. Política de despacho dos servidores.

Essa hipótese determina que seja feita uma lista de preferências de despacho de servidores para cada átomo. No caso do primeiro servidor da lista estar ocupado no momento de um chamado, despacha-se o próximo servidor, caso este também esteja ocupado, despacha-se o próximo da lista e assim sucessivamente. Caso todos estejam ocupados, o chamado será perdido, se o sistema não admitir fila, ou entrará numa fila. Essa lista deverá permanecer inalterada durante todo o funcionamento do sistema.

8. Composição do tempo total de atendimento.

O tempo total de atendimento é considerado desde o telefonema solicitante até o retorno da equipe à base. É composto pelo somatório dos seguintes tempos: tempo de preparo do servidor (*setup time*), tempo de viagem do servidor até o local da ocorrência, tempo de execução do serviço junto ao usuário (tempo em cena) e o tempo de retorno à base. As taxas médias de atendimento μ_n de cada servidor poderão ser diferentes. O modelo hipercubo é mais eficiente, no caso do sistema admitir fila de espera, se os tempos de atendimento se aproximarem dos desvios-padrões. Porém, segundo Larson e Odoni (1981), desvios razoáveis dessa hipótese não alteram a precisão do modelo. Essa hipótese torna-se menos importante se o sistema não admitir filas de espera.

9. Dependência de atendimento em relação ao tempo de viagem:

É necessário verificar se os tempos médios de viagem são pequenos em relação ao tempo de atendimento (junto ao paciente). Essa condição é normalmente atendida em áreas urbanas e menos atendida nas áreas rurais, e em serviços prestados em rodovias (LARSON e ODONI, 1981).

Na prática, poucos são os sistemas reais que aderem completamente a todas as hipóteses mencionadas. Portanto, em cada caso deve ser feita uma análise da complexidade do problema em questão para verificar a aplicabilidade do modelo (LARSON e ODONI, 1981).

3.3.2 Transições entre estados e equações de equilíbrio

O modelo hipercubo baseia-se na resolução de um sistema de equações lineares. São essas equações que fornecem as probabilidades de equilíbrio dos possíveis estados do sistema e as medidas de desempenho. Porém, com o aumento do número de servidores há um aumento exponencial de estados do modelo. O cálculo das medidas de desempenho para um sistema que opera com N servidores e não permite filas de espera, envolve uma solução de um sistema de 2^N equações lineares. Quanto maior o número de servidores, maior o número de equações podendo tornar inviável a solução computacional em relação ao tempo.

A fim de aplicar o modelo, a probabilidade de que cheguem ao sistema dois chamados simultaneamente é considerada nula. Também, é considerada nula a probabilidade de dois servidores estarem livres simultaneamente. Com isso, em qualquer transição é permitido apenas um passo por vez. As taxas de transições de um estado para outro são chamadas de ascendentes ou descendentes.

A solução do modelo é dada pela construção das equações de equilíbrio do sistema, que são definidas supondo-se que o sistema atinja o equilíbrio. Assim, o fluxo que entra neste estado deve ser igual ao fluxo que sai deste estado. Em um sistema não saturado, com capacidade de fila infinita, as probabilidades de estado do modelo hipercubo são calculadas a partir das equações de balanço, construídas a partir dos possíveis estados.

Será apresentado a seguir um exemplo ilustrativo, resolvido por Chiyoshi *et al* (2001), com apenas três servidores e três átomos, a fim de facilitar o entendimento das taxas de transição de estados (λ e μ), veja a (FIGURA 3.7).

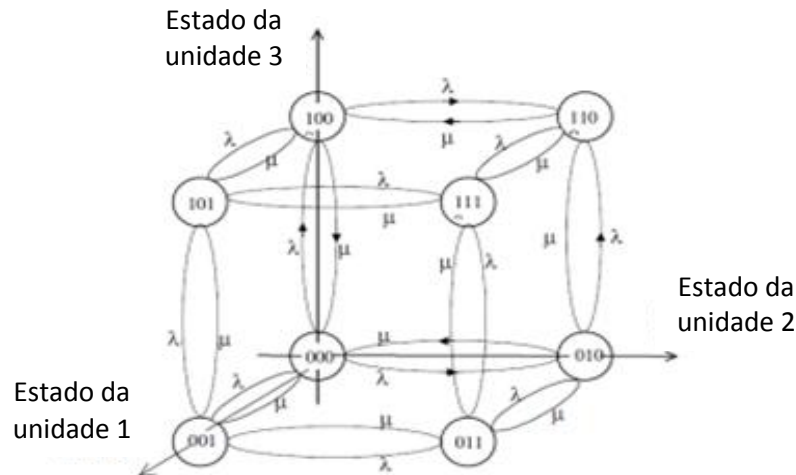


FIGURA 3.7 - TAXA DE TRANSIÇÃO DE ESTADO ENTRE OS SERVIDORES
FONTE: Costa (2003)

Nesse caso, conforme mencionado anteriormente, tem-se oito possíveis estados em que o sistema pode se encontrar num determinado instante. São eles:

- Estado $S'_0 = (000)$ (todos os servidores encontram-se desocupados);
- Estados $S'_1 = (001)$ ou (010) ou (100) (apenas um servidor ocupado);
- Estados $S'_2 = (110)$ ou (101) ou (011) (dois servidores ocupados);
- Estado $S'_3 = (111)$ (todos os servidores ocupados).

Se considerarmos para o exemplo ilustrativo fila de espera, tem-se ainda a possibilidade de existir outros estados (S'_4, S'_5, \dots, S'_n), sendo que se existir um S'_i tal que $i \geq 4$, isso indicará que três servidores estão ocupados e que existem $i - 3$ chamadas numa fila de espera.

A transição de estados no modelo hipercubo de filas ocorre de maneira análoga aos modelos clássicos de filas.

O modelo hipercubo de filas encontra soluções a partir da construção das equações de equilíbrio, nos quais os resultados baseiam-se nos valores das probabilidades de cada estado dos modelos clássicos de filas. As equações de equilíbrio são modeladas partindo do pressuposto de equilíbrio no sistema (*steady state*), ou seja, para cada estado do sistema, o fluxo que chega neste estado deverá ser igual ao fluxo que sai dele.

Considerando o exemplo ilustrativo (CHIYOSHI ET AL, 2001), com oito equações de equilíbrio, sem considerar fila de espera. Podem ocorrer as seguintes situações:

1ª Situação: Todos os servidores encontram-se livres, estado $S_0 = (000)$

Nesse caso, em que todos os servidores encontram-se livres, existem três possibilidades de ocorrência. O primeiro servidor passa a ocupado $S_1 = (001)$, com uma taxa de ocorrência λ_1 , ou o segundo servidor passa para o estado ocupado $S_2 = (010)$, com uma taxa de ocorrência λ_2 , ou o terceiro servidor passa a ocupado $S_3 = (100)$ com uma taxa de ocorrência λ_3 . A taxa total de transição ascendente para sair do estado livre e passar para o estado ocupado será a soma de todas as taxas $\lambda_1 + \lambda_2 + \lambda_3 = \lambda$. Lembrando que apenas um passo é permitido por vez, não existe a possibilidade de dois servidores ficarem ocupados ao mesmo tempo.

De maneira análoga ocorre o inverso. A transição descendente para chegar-se ao estado vazio $S_0 = (000)$, ocorre ou se o primeiro servidor $S_1 = (001)$ terminou seu atendimento, com uma taxa μ_1 , ou se o segundo servidor $S_2 = (010)$ terminou seu atendimento, com uma taxa μ_2 , ou se o terceiro servidor $S_3 = (100)$ terminou seu atendimento, com uma taxa μ_3 .

Considerando-se que apenas um passo pode acontecer de um estado para outro, o modelo não admite que dois servidores fiquem livres simultaneamente.

Dessa forma, para o estado livre $S_0 = (000)$, a primeira equação de equilíbrio será igual a:

$$\lambda \cdot p_{000} = \mu_1 \cdot p_{001} + \mu_2 \cdot p_{010} + \mu_3 \cdot p_{100} \quad (3.35)$$

Sendo:

P_B – probabilidade de equilíbrio do estado B é $\lambda = \lambda_1 + \lambda_2 + \lambda_3$

2ª Situação: Apenas um servidor encontra-se ocupado, neste caso existem três possibilidades, estados $S_1 = (001)$, $S_2 = (010)$ e $S_3 = (100)$

Inicialmente analisa-se o estado $S_1 = (001)$.

Estando o sistema com o primeiro servidor ocupado, a chegada de um novo chamado gera, neste caso, dois servidores ocupados, por exemplo, (101) ou (011) . Também poderá ocorrer do único servidor ocupado terminar seu atendimento, no caso passará do estado $S_1 = (001)$ para o estado livre $S_0 = (000)$. Com isso, a taxa

de transição do estado $S_1 = (001)$ para os outros possíveis estados será a soma de $\lambda + \mu_1$.

No sentido inverso, para alcançar o estado $S_1 = (001)$ tem-se as seguintes possibilidades:

- O sistema encontra-se no estado $S_0 = (000)$, todos servidores livres e chega um chamado oriundo do átomo 1;
- O sistema encontra-se no estado $S_4 = (101)$ e o terceiro servidor conclui seu atendimento;
- O sistema encontra-se no estado $S_5 = (011)$ e o segundo servidor conclui seu atendimento.

Nesse caso a equação de equilíbrio para o estado $S_1 = (001)$ será igual:

$$(\lambda + \mu_1) \cdot p_{001} = \lambda_1 \cdot p_{000} + \mu_2 \cdot p_{011} + \mu_3 \cdot p_{101} \quad (3.36)$$

Com o mesmo raciocínio encontram-se as equações de equilíbrio para os estados $S_2 = (010)$ e $S_3 = (100)$, respectivamente:

$$(\lambda + \mu_2) \cdot p_{010} = \lambda_2 \cdot p_{000} + \mu_1 \cdot p_{011} + \mu_3 \cdot p_{110} \quad (3.37)$$

$$(\lambda + \mu_3) \cdot p_{100} = \lambda_3 \cdot p_{000} + \mu_1 \cdot p_{101} + \mu_2 \cdot p_{110} \quad (3.38)$$

3ª Situação: Apenas um servidor encontra-se livre, estados $S_4 = (101)$, $S_5 = (011)$ e $S_6 = (110)$

Analisando inicialmente o estado $S_4 = (101)$.

A chegada de um chamado no sistema levará para o estado $S_7 = (111)$, todos os servidores ocupados, mas também existe a possibilidade de o primeiro servidor desocupar, passando para o estado $S_3 = (100)$, ou ainda o terceiro servidor desocupar, passando para o estado $S_1 = (001)$. Com isso, a taxa total de transição do estado $S_4 = (101)$, para os outros estados será igual a soma de:

$$\lambda + \mu_3 + \mu_1 \quad (3.39)$$

No sentido inverso, para alcançar o estado $S_4 = (101)$ tem-se as seguintes possibilidades:

- O sistema encontra-se no estado $S_1 = (001)$ com primeiro servidor está ocupado, e chega um chamado oriundo do átomo 3;
- O sistema encontra-se no estado $S_3 = (100)$, o terceiro servidor está ocupado, e chega um chamado oriundo do átomo 1;
- O sistema encontra-se no estado $S_7 = (111)$ e o segundo servidor conclui seu atendimento.

Nesse caso a equação de equilíbrio para o estado $S_4 = (101)$ será igual a:

$$(\lambda + \mu_3 + \mu_1) \cdot p_{101} = \lambda_3 \cdot p_{001} + (\lambda_1 + \lambda_3) \cdot p_{100} + \mu_2 \cdot p_{111} \quad (3.40)$$

Com o mesmo raciocínio encontram-se as equações de equilíbrio para os estados $S_5 = (011)$ e $S_6 = (110)$, respectivamente:

$$(\lambda + \mu_2 + \mu_1) \cdot p_{011} = (\lambda_1 + \lambda_2) \cdot p_{001} + \lambda_1 \cdot p_{010} + \mu_2 \cdot p_{111} \quad (3.41)$$

$$(\lambda + \mu_2 + \mu_3) \cdot p_{110} = (\lambda_3 + \lambda_2) \cdot p_{010} + \lambda_2 \cdot p_{100} + \mu_1 \cdot p_{111} \quad (3.42)$$

4ª Situação: Todos os servidores estão ocupados, estado $S_7 = (111)$

Nesse estado $S_7 = (111)$, onde todos os servidores encontram-se ocupados, e supondo não haver usuário na fila de espera, qualquer chegada de chamado ou término de atendimento provocará uma transição do sistema para outro estado. Para alcançar esse estado $S_7 = (111)$ basta que dois servidores estejam atendendo e chegue mais um chamado. O servidor que se encontrar livre será automaticamente despachado para realizar o atendimento, sendo ele preferencial ou *backup*. Também é possível alcançar o estado $S_7 = (111)$ a partir de um estado S_8 , quando todos os servidores estão atendendo e têm um usuário na fila de espera, no caso do sistema admitir fila de espera, assim que um servidor termine seu atendimento, o usuário da fila será atendido. Neste caso, consideremos p_8 a probabilidade de se encontrar no estado S_8 , ou seja, $p_8 = P(S_8)$. Com isso, a equação de equilíbrio do estado $S_7 = (111)$ será dada por:

$$(\lambda + \mu).p_{111} = \lambda.p_{011} + \lambda.p_{101} + \lambda.p_{110} + \mu.p_8 \quad (3.43)$$

sendo:

$$\mu = \mu_1 + \mu_2 + \mu_3$$

Esse procedimento poderia ser repetido infinitamente para outros estados, S_9, S_{10}, \dots , entretanto para respeitar as condições de equilíbrio do sistema, as taxas de transições devem ser iguais a $\lambda.p_{111} = \mu.p_8$. Por outro lado, se $\lambda.p_{111} > \mu.p_8$, a fila, também conhecida como cauda, aumentaria infinitamente e o sistema entraria em um estado transiente. Caso contrário, se $\lambda.p_{111} < \mu.p_8$, o modelo estaria em crescimento em termos da massa de probabilidade (CHIYOSHI *ET AL*, 2000).

Dessa forma, a equação de equilíbrio do estado p_{111} , poderá ser descrita como:

$$\mu.p_{111} = \lambda.p_{011} + \lambda.p_{101} + \lambda.p_{110} \quad (3.44)$$

Com isso obtêm-se, para o exemplo simples, sem fila de espera, oito equações de equilíbrio, que formam um sistema linear com oito incógnitas, a citar: $p_{000}, p_{001}, p_{010}, p_{100}, p_{011}, p_{101}, p_{110}$ e p_{111} . Segundo Chiyoshi *et al* (2000), tem-se então um sistema linear homogêneo, possível e indeterminado na forma $Ax = 0$. Isso se dá devido ao fato das equações fornecerem apenas o equilíbrio, no qual impõem condições de equilíbrio para cada estado do sistema, porém nada fornecem quanto à forma como a massa de probabilidade se distribui entre os estados. Dessa maneira, a fim de eliminar essa indeterminação, introduz-se uma equação de normalização, considerando-se que a soma das probabilidades de todos os possíveis estados do sistema deve ser igual a um (CHIYOSHI *et al*, 2000), (ALBINO, 1994) tem-se:

$$p_{000} + p_{001} + p_{010} + p_{100} + p_{011} + p_{101} + p_{110} + p_{111} = 1 \quad (3.45)$$

Nesse caso, substituindo a equação 3.45 por qualquer uma das equações de equilíbrio anteriores (3.35, 3.36, 3.37, 3.38, 3.40, 3.41, 3.42 e 3.44), tem-se um sistema possível e determinado, e a partir daí encontram-se os valores de todas as probabilidades de equilíbrio de estado.

Existem outros tipos de sistemas representativos. O sistema pode admitir fila de espera (cauda) indeterminada e a substituição da equação de normalização do modelo hipercubo de filas é dado por:

$$p_{000} + p_{001} + p_{010} + p_{100} + p_{011} + p_{101} + p_{110} + p_{111} = 1 - p_Q \quad (3.46)$$

$$\text{com } p_Q = 1 - \sum_{n=0}^N p_n$$

p_Q é a probabilidade de haver fila de comprimento positivo;

p_n é a probabilidade de haver n servidores ocupados.

E podem-se ter ainda, sistemas que limitam o número de usuários aguardando em fila de espera. Nesse caso, o modelo hipercubo de fila seria com cauda limitada (TAKEDA, 2000).

3.4 MEDIDAS DE DESEMPENHO DO SISTEMA

A partir das equações de equilíbrio do sistema se obtém as probabilidades dos estados e a partir destas probabilidades calculam-se as medidas de desempenho listadas a seguir. Essas medidas de desempenho auxiliam a análise sob a hipótese de que o sistema esteja em equilíbrio, podendo expressar o desempenho de um determinado servidor para um determinado átomo, ou ainda para o sistema como um todo. Os servidores são considerados distintos, com tempos de atendimentos diferentes.

- I. Carga de trabalho ou *workload* (ρ): é a carga de trabalho de um servidor i , corresponde à fração do tempo em que o servidor encontra-se ocupado, ou seja, a soma das probabilidades de estado do servidor encontrar-se em atendimento.

$$\rho_i = \sum_{B: b_i=1} p_B + p_Q \quad (3.47)$$

Em que:

- ρ_i é a carga de trabalho do servidor i ($i = 1, 2, \dots, n$);
- $\sum_{B:b_i=1} p_B$ é soma das probabilidades dos estados do servidor i encontrar-se ocupado;
- p_Q é a probabilidade de haver fila de espera.

Para o exemplo ilustrativo, com três servidores, tem-se:

$$\rho_1 = p_{001} + p_{011} + p_{101} + p_{111} + p_Q$$

$$\rho_2 = p_{010} + p_{011} + p_{110} + p_{111} + p_Q$$

$$\rho_3 = p_{100} + p_{110} + p_{101} + p_{111} + p_Q$$

E para um sistema infinito, um servidor n , terá seu workload (ρ_n) escrito como:

$$\rho_n = p_{10\dots0} + p_{10\dots01} + \dots + p_{1\dots11} + p_Q$$

II. Frequências de despachos

Frequência de despacho é outra medida muito utilizada nos SAE móvel. É a medida da fração de despacho de um determinado servidor i até um átomo j (f_{ij}), dentro do sistema. É uma medida composta de duas parcelas, a primeira $f_{ij}^{[1]}$ está relacionada à fração dos despachos em que o servidor i é enviado até o átomo j , sem levar em consideração tempo de espera em fila, a segunda parcela $f_{ij}^{[2]}$ está relacionada também à fração dos despachos do servidor i até o átomo j , porém levando em consideração tempo de espera na fila.

$$f_{ij} = f_{ij}^{[1]} + f_{ij}^{[2]} \quad (3.48)$$

$$f_{ij}^{[1]} = \frac{\lambda_j}{\lambda} \sum_{B \in E_{ij}} p_B \quad (3.49)$$

$$f_{ij}^{[2]} = \frac{\lambda_j}{\lambda} p_s \frac{\mu_i}{\mu} \quad (3.50)$$

$$f_{ij} = \frac{\lambda_j}{\lambda} \sum_{B \in E_{ij}} P_B + \frac{\lambda_j}{\lambda} P_S \frac{\mu_i}{\mu} \quad (3.51)$$

No qual:

$\frac{\lambda_j}{\lambda} \rightarrow$ probabilidade do servidor i ser o primeiro disponível na lista de preferência de despacho do átomo j ;

$E_{ij} \rightarrow$ conjunto de todos os estados nos quais o servidor i é o primeiro servidor disponível na lista de preferência de despacho ao átomo j ;

$\frac{\mu_i}{\mu} \rightarrow$ probabilidade do servidor i ser o primeiro a ficar livre, uma vez que todos os servidores estão ocupados (fila);

P_S é o estado de saturação do sistema.

Tanto a primeira parcela da equação ($f_{ij}^{[1]}$) quanto a segunda ($f_{ij}^{[2]}$) referem-se à probabilidade de surgir um novo chamado ao átomo j antes dos demais, uma vez que os chamados chegam a partir de um processo de Poisson. No entanto, na segunda parcela, poderá ter fila de espera. Resumindo, o valor de ($f_{ij}^{[2]}$) é o produto de três termos. O primeiro é a probabilidade condicional de que o chamado seja proveniente do átomo j , o segundo é a probabilidade de que uma chegada aleatória de um chamado fique na fila de espera e o terceiro é a probabilidade condicional de que um servidor i seja despachado.

Com relação a essas frequências de despachos podem-se calcular outras importantes medidas de desempenho. (LARSON E ODONI, 1981).

II.1) Fração dentre todos os despachos que são interáreas de cobertura:

$$f_I = \sum_{n=1}^N \sum_{j \notin \text{área de cobertura primária de } n} f_{nj} \quad (3.52)$$

II.2) Fração dos despachos do servidor n que são interárea de cobertura:

$$f_{In} = \frac{\sum_{j \notin \text{área de cobertura primária de } n} f_{nj}}{\sum_{j=1}^{N_A} f_{nj}} \quad (3.53)$$

II.3) Fração dos chamados da área de cobertura i que são atendidas por outro servidor que não seja i :

$$f'_{ii} = \frac{\sum_{n \neq i} \frac{\sum_{j \in \text{área de cobertura primária de } i} f_{nj}}{N}}{\sum_{n=1} \sum_{j \in \text{área de cobertura primária de } i} f_{nj}} \quad (3.54)$$

III. Tempo médio de deslocamento no sistema

Esses valores são encontrados a partir de uma matriz de origem-destino entre todos os pares de átomos (τ_{ij}) , obtidas no próprio sistema. Caso não seja possível medir esses valores, deve-se estimá-los utilizando alguns conceitos de probabilidade geométrica (LARSON e ODONI, 1981). Os tempos gastos para ir de um átomo i a um átomo j poderá ser diferente do tempo gasto no caminho inverso. Isso se deve a vários fatores, entre eles, o trânsito, horário, dia da semana, entre outros.

Alguns tempos médios de deslocamentos:

a. Tempos médios de viagem no sistema:

A fim de encontrar os tempos médios de deslocamento no sistema, faz-se necessário conhecer a localização dos servidores no momento em que este é despachado para atender a um chamado, o tempo médio de um servidor i , quando livre, deslocar-se até o local do chamado e também o tempo de espera de um chamado que se encontra numa fila.

A representação dessas localizações se dá sob a forma de matriz $L = [l_{ij}]$, na qual cada elemento da matriz representa a probabilidade do servidor se encontrar num determinado átomo quando disponível. A matriz L é uma matriz estocástica, isto é,

$$\sum_{j=1}^{N_A} l_{ij} = 1$$

em que:

$l_{ij} = 1$ se o servidor encontra-se no átomo j , e $l_{ij} = 0$ caso contrário.

O tempo médio para um servidor deslocar-se até um determinado átomo é dado por:

$$t_{ij} = \sum_{k=1}^{N_A} l_{ik} \cdot \tau_{kj} \quad i = 1, 2, \dots, N \quad e \quad j = 1, 2, \dots, N_A \quad (3.55)$$

No qual:

τ_{kj} representa a matriz de tempo de deslocamento entre os átomos k e j .

O tempo médio de espera para chamados aguardando em filas é calculado por:

$$\bar{T}_Q = \sum_{i=1}^{N_A} \sum_{j=1}^{N_A} \frac{\lambda_i \lambda_j}{\lambda^2} \cdot \tau_{ij} \quad (3.56)$$

Do qual:

$\frac{\lambda_j}{\lambda} \rightarrow$ probabilidade de um chamado que está na fila ter ocorrido no átomo j ;

$\frac{\lambda_i}{\lambda} \rightarrow$ probabilidade desse chamado ter sido atendido por um servidor localizado no átomo i , respectivamente;

Considerando a probabilidade de saturação do sistema (P_S), o tempo médio de deslocamento nesse caso será obtido por:

$$\bar{T} = \sum_{i=1}^N \sum_{j=1}^{N_A} f_{ij}^{[1]} \cdot t_{ij} + P_S \cdot \bar{T}_Q \quad (3.57)$$

b. Tempos médios de viagem para cada átomo:

Essa medida reflete a qualidade do serviço oferecido pelo sistema. Com isso, torna-se uma medida muito importante, pois é o tempo de viagem para cada

átomo j (\bar{T}_j). Também conhecido como Tempo Médio de Resposta (TMR), nos sistemas reais.

$$\bar{T}_j = \frac{\sum_{n=1}^N f_{nj}^{[1]} \cdot t_{nj}}{\sum f_{nj}^{[1]}} (1 - P_s) + \sum_{i=1}^{N_A} \frac{\lambda_i}{\lambda} \cdot \tau_{ij} \cdot P_s \quad (3.58)$$

Da mesma forma pode-se calcular o tempo médio de viagem para cada servidor n :

$$\overline{TU}_n = \frac{\sum_{j=1}^{N_A} f_{nj}^{[1]} \cdot t_{nj} + \left(\frac{\bar{T}_Q \cdot P_s}{N} \right)}{\sum_{j=1}^{N_A} f_{nj}^{[1]} + \left(\frac{P_s}{N} \right)} \quad (3.59)$$

Pode-se também calcular o tempo médio de deslocamento até os átomos da área de cobertura primária n :

$$\overline{TRA}_n = \frac{\sum_{j \in \text{áreadecoberturan}} \sum_{m=1}^N f_{mj}^{[1]} \cdot t_{mj}}{\sum_{j \in \text{áreadecoberturan}} \sum_{m=1}^N f_{mj}^{[1]}} (1 - P_s) + \frac{\sum_{k \in \text{áreadecoberturan}} \sum_{j=1}^{N_A} \frac{\lambda_j \cdot \lambda_k \cdot \tau_{jk}}{\lambda^2}}{\sum_{k \in \text{áreadecoberturan}} \frac{\lambda k}{\lambda}} \cdot P_s \quad (3.60)$$

Segundo Larson e Odoni (1981, 2007), o modelo permite calcular outras medidas de desempenho. Para isso, faz-se necessário verificar a aplicabilidade no problema abordado.

A resolução do modelo hipercubo de filas consiste em encontrar a solução de um sistema de equações de equilíbrio linear. Conforme já mencionado, o número de equações está diretamente ligado ao número de servidores (veículos) do sistema. Como o número de equações é dado 2^n , o aumento do número n provoca um aumento exponencial do número de equações e, com isso, o tempo computacional poderá tornar-se inviável. Dessa forma, deve-se analisar inicialmente o número de servidores (n) do problema abordado, assegurar a viabilidade computacional. A

resolução do problema por métodos exatos, em geral, torna-se inviável. Por esse motivo, deve-se avaliar o uso de métodos heurísticos para a resolução do sistema, tais como o método de Larson (1972), que envolve um sistema não linear de equações, e tem como incógnitas as m taxas de ocupação dos servidores (*workloads* dos servidores), e não as probabilidades de estado descritas pelos vértices do hipercubo, usadas nos métodos exatos.

3.5 ALGORITMO ANT COLONY OPTIMIZATION

O algoritmo *Ant Colony Optimization* (ACO), é uma meta-heurística inspirada no comportamento forrageiro das formigas em busca de alimento. O termo inspiração refere-se à organização do trabalho e cooperação entre si. A auto-organização das formigas em atividades coletivas originou o interesse em se criar sistemas artificiais inteligentes baseados nessas características. São diversos os casos em que a natureza serviu de inspiração para a criação de sistemas artificiais. Pode-se citar como exemplos: nuvem de partículas, inspirada no vôo dos pássaros; redes neurais artificiais, inspiradas no neurônio biológico; algoritmo genético; entre outros.

No caso das formigas, um fator interessante é que muitas espécies são praticamente cegas. Nesse caso, a comunicação entre as formigas é realizada de duas formas: a primeira forma é a direta, contato entre si via mandíbulas, antenas, ou contato químico. Na segunda, essa interação pode ocorrer de maneira indireta, a partir do momento em que dois indivíduos interagem, e um deles muda o ambiente e o outro reage a essa mudança posteriormente (BONABEAU *et al.*, 1999). Essa mudança no ambiente é conhecida segundo Dorigo e Stutzle (2004), como *stigmergy* e quer dizer estímulo dos trabalhadores para melhorar o desempenho. No caso das formigas, elas percebem essa mudança no ambiente, realizada por suas companheiras de maneira indireta, e este se torna seu meio de comunicação. A aplicabilidade do funcionamento das colônias de formigas em um algoritmo artificial está fundamentada nessa comunicação indireta, que pode ser operacional e flexível.

Basicamente, a comunicação entre as formigas é realizada por meio de uma substância química produzida por elas, denominada feromônio. Ao caminhar, as formigas depositam no chão o feromônio, formando uma trilha. As demais formigas sentem o cheiro da substância, e quando elas têm que escolher um caminho, elas

escolhem o caminho com maior quantidade de feromônio. Essas trilhas de feromônios são utilizadas para marcar o caminho do ninho até o alimento disponível, bem como o caminho de volta.

Um experimento muito interessante, conhecido como experimento da ponte binária, realizado por Deneubourg *et al.* (1990), consiste em uma ponte dupla com formigas em busca de alimentos, conforme mostra a (FIGURA 3.8).

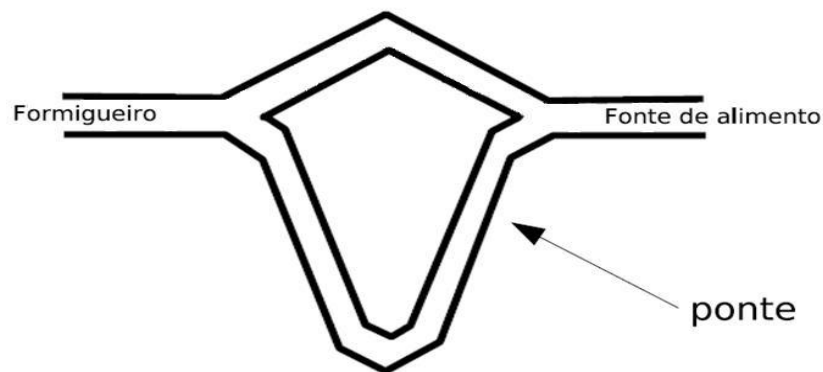


FIGURA 3.8 - PONTE BINÁRIA BASEADO EM GOSS *ET AL.* (1989)

Inicialmente, as formigas são livres para escolher seus caminhos, ainda sem feromônio e realizam suas buscas por possíveis soluções aleatoriamente. As formigas convergem para um dos caminhos com igual probabilidade. Ao encontrarem alimento, retornam ao formigueiro depositando feromônio na trilha. A trilha que contém maior quantidade desta substância química indica que um maior número de formigas passou por ali, logo maior será a probabilidade desse caminho ser o melhor e mais curto. Relacionando esse comportamento forrageiro com o algoritmo, esse caminho encontrado representa uma solução otimizada que é função da quantidade de substância química ali depositada.

Os resultados dos experimentos mostraram que as formigas convergiram para a ponte mais curta, aquela com maior quantidade de feromônio. Quanto mais elas depositavam o feromônio, maior era o estímulo nas outras formigas em seguir a mesma trilha, convergindo assim para o mesmo caminho.

O algoritmo colônia de formigas foi criado na década de 1990, Dorigo (1992). O sistema artificial e a análise probabilística do comportamento foram totalmente inspirados no modelo natural da trilha de feromônio das formigas. As formigas artificiais também são conhecidas na literatura como agentes.

A formiga tende a intensificar sua busca por alimentos no caminho com maior quantidade de feromônio, o que lhe proporciona um caminho mais curto e conseqüentemente um menor tempo de busca pela solução.

As formigas reais interagem basicamente pelo feromônio, e é essa interação que é implementada na meta-heurística de colônia de formigas artificiais para resolver problemas de otimização. O algoritmo ACO trata de uma meta-heurística construtiva que constrói soluções de forma probabilística utilizando algumas informações. Uma destas informações é a trilha de feromônio artificial que muda dinamicamente durante a execução do programa de modo a refletir a experiência já adquirida durante a busca. Existe também a retroalimentação positiva em função das trilhas, que trata do acréscimo de feromônio artificial nas boas soluções encontradas e do decréscimo nas soluções ruins, o que caracteriza o comportamento cooperativo entre as formigas, buscando atingir melhor desempenho.

A (TABELA 3.1) representa resumidamente a correlação entre o que acontece na natureza e no algoritmo ACO.

TABELA 3.1 - CORRESPONDÊNCIA ENTRE A NATUREZA E O ACO

Natureza	ACO
Possíveis caminhos entre o formigueiro (ninho) e o alimento	Conjunto de possíveis soluções (diferentes vetores das variáveis)
Caminho mais curto	Melhor Solução
Ação via comunicação por feromônio	Procedimento de otimização

FONTE: KOIDE (2010)

A representação de grafos é muito utilizada no algoritmo colônia de formigas. Um grafo é representado por $G(N,A)$, no qual N é o conjunto de nós e A representa as arestas entre os nós. A solução do algoritmo consiste na seleção probabilística dos nós selecionados. As formigas possuem "registros" das trilhas já percorridas. Dessa forma, o algoritmo terá as seguintes características:

- Construir solução probabilística para a trilha do feromônio sem atualização durante o percurso;
- Atualizar a matriz de feromônio no caminho inverso;

- Avaliar a qualidade das soluções obtidas e determinar a nova quantidade de feromônio a depositar ou não.

3.5.1 Problema do caixeiro viajante

Um dos problemas que podem ser resolvidos com o ACO é o Problema do Caixeiro Viajante (PCV). O PCV é um problema clássico da Pesquisa Operacional (PO) e de larga aplicabilidade, muito estudado na literatura por diversos pesquisadores. Segundo Karp (1975), é um problema simples de ser formulado, contudo difícil de ser resolvido, uma vez que é *NP-hard*. O PCV pode ser resolvido por métodos exatos ou métodos heurísticos. Contudo, os métodos exatos apresentam uma grande complexidade fatorial, aumentando muito o tempo computacional quanto maior for a lista de cidades. Já os métodos heurísticos, apesar de não apresentarem soluções exatas, apresentam soluções aproximadas satisfatórias e com baixo tempo computacional, quando comparados ao método exato. Algumas das técnicas utilizadas para a resolução do PCV são: Algoritmos genéticos, GRASP (*Greedy Randomized Adaptive Search Procedures*) e ACO.

O PCV consiste em encontrar o percurso de um caixeiro que, partindo de uma cidade qualquer deve visitar a próxima cidade, de uma lista de preferências, uma única vez, e retornar a cidade de onde partiu. O ciclo total percorrido deve ter o menor custo (c_{ij}). Assim, o PCV basicamente consiste em encontrar um circuito de menor custo. Cada cidade está associada a um vértice do grafo. As distâncias entre as cidades são arestas valoradas (i, j) do grafo $G(N, A)$, onde $N = \{1, \dots, n\}$ é o conjunto de nós e $A = \{1, \dots, m\}$ é o conjunto de arestas de G . A aplicação desse problema se encaixa nos mais diversos ramos da literatura, por exemplo: minimização de rotas de veículos, confecção de sistemas digitais, sequenciamento de atividades, dentre outros.

O PCV pode ser classificado como simétrico, caso todos os pares de nós (i, j) tenham os mesmos custos, ou assimétrico quando os custos forem diferentes. Hoffman e Wolfe (1985) apresentam uma revisão mais completa sobre o PCV.

3.5.2 ACO Aplicada ao Problema do Caixeiro Viajante

O problema de um caixeiro viajante foi enunciado em 1934 pelo matemático Karl Menger. Foi ele quem primeiro apresentou o problema sob o ponto de vista matemático. O PCV foi o primeiro problema de roteamento apresentado na literatura, e até os dias atuais não se conhece nenhum algoritmo computacional que o resolva em tempo eficiente. Por isso, até os dias de hoje, muitos pesquisadores apresentam grande interesse na busca de solução para o problema.

O algoritmo ACO desenvolvido inicialmente por Dorigo *et al* (1996) e Dorigo e Gambardella (1997) tinha como principal objetivo resolver o PCV simétrico e assimétrico. Contudo, a eficiência e robustez do ACO, foram comprovadas com sua aplicação em problemas *job-shop scheduling* e roteamento de veículos.

A solução do PCV clássico pelo algoritmo ACO é relativamente simples. Inicialmente, cada formiga parte de uma cidade escolhida aleatoriamente, ou seja, de um vértice do grafo, a escolha da próxima cidade que fará parte da rota é probabilística, e está diretamente associada ao valor do feromônio e ao valor da atratividade ($1/d_{ij}$) que por sua vez leva em conta a distância (d_{ij}) entre as cidades i e j . A quantidade inicial de feromônio é igual e constante para todas as arestas. O caminho de cada formiga é guardado em sua memória evitando dessa forma, a repetição de cidades. A construção da solução termina quando a formiga passa por todos os vértices do grafo. Neste momento, calcula-se o valor das soluções. Assim que cada formiga completa seu circuito, a matriz de feromônio é atualizada. A quantidade de feromônio acrescida a cada aresta está relacionada à qualidade da solução encontrada. Quanto melhor a solução encontrada, maior será a quantidade de feromônio depositada na referida aresta. O processo continua até que seja satisfeito algum critério de parada.

O algoritmo do ACO é baseado no cálculo das probabilidades p_{ij}^k , que define a probabilidade da formiga k ir da cidade i para a cidade j . Segundo Dorigo e Stutzle (2004), esta probabilidade é calculada conforme equação 3.61.

$$p_{ij}^k = \begin{cases} \frac{[\tau_{ij}]^\alpha [n_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}]^\alpha [n_{il}]^\beta}, & j \in N_i^k \\ 0 & j \notin N_i^k \end{cases} \quad (3.61)$$

Sendo que:

τ_{ij} representa a quantidade de feromônio da aresta ij ;

α parâmetro que controla a influência de feromônio;

η_{il} valor heurístico definido em função das características do problema, no caso do PCV é dado por $1/d_{ij}$;

β parâmetro que controla a influência da informação heurística

I candidato do conjunto de soluções;

N_i^k conjunto de vértices de vizinhança variável do vértice i , associado à k -ésima formiga;

k Índice (identificador único) da formiga.

A atualização da quantidade de feromônio é estabelecida de acordo com a equação 3.62 a seguir.

$$\tau_{ij}^k \leftarrow (1 - \rho)\tau_{ij}^k + \sum \Delta\tau_{ij}, \quad \forall(i, j) \in A; \quad \rho \in [0,1] ; \quad \Delta\tau_{ij} = 1/L_k \quad (3.62)$$

Nesse caso o ρ representa um parâmetro da taxa de fixação do feromônio, que deve ser um valor entre zero e um, para evitar que o acúmulo de feromônio entre duas cidades quaisquer deixe o algoritmo ineficiente. O valor inicial para todas as formigas é igual a $\Delta\tau^k$ constante. Devido à variação de feromônio, caminhos melhores e mais curtos são alcançados em razão da intensificação da substância.

Em outros casos a quantidade de feromônio pode diminuir no decorrer do tempo. Com isso, novos caminhos poderão ser explorados durante o processo de busca. Há um limite mínimo para a evaporação do feromônio, evitando assim uma estagnação da solução em um ótimo local.

O (ALGORITMO 3.1) ilustra os principais passos do procedimento ACO.

Coloque cada formiga em uma cidade aleatória

```

para  $t = 1$  até número de iterações
  para  $k = 1$  até número de formigas
    enquanto a formiga  $k$  não construir a viagem  $S_k$ 
      Selecione a próxima cidade pela regra  $p_{ij}^k$ 
    fim enquanto
    Calcule a distância  $L_k$  da viagem  $S_k$ 
    se  $L_k < L^*$  então
       $S^* = S_k$ ,  $L^* = L_k$ 
    fim se
  fim para
  Atualize os feromônios
fim para
retornar  $S^*$ 

```

ALGORITMO 3.1 - PSEUDO-CÓDIGO DO ALGORITMO ACO.

FONTE: SILVA (2005)

3.5.3 ACO para o problema de múltiplos caixeiros viajantes (PMCV)

Este estudo propõe uma mudança no algoritmo ACO, a fim de aplicá-lo ao Problema de Múltiplos Caixeiros Viajantes (PMCV). Cada caixeiro irá definir um conjunto de pontos do espaço geográfico (rotas), e a esses conjuntos de pontos denominam-se átomos geográficos. O objetivo é fazer a adaptação do ACO com o intuito de obter novas divisões geográficas. Dessa forma, aplicando o modelo hipercubo de filas num cenário diferente, com novas divisões geográficas, tem-se novas medidas de desempenho do SAE móvel. Com isso, melhor será a avaliação do SAE por meio das diversas medidas de desempenho fornecidas pelo modelo hipercubo.

O PMCV é uma extensão do PCV diferindo deste pelo fato de determinar múltiplas rotas. O objetivo é diminuir o custo total, de modo que cada caixeiro viajante deva visitar pelo menos um nó do grafo e cada nó só pode ser visitado uma única vez.

O PRV é uma generalização do PCV, em que as cidades são substituídas por clientes e os caixeiros por veículos, com capacidade limitada. Segundo Mazzeo e Loiseau (2004), as principais características da otimização por Colônia de

Formigas para o Problema de Roteamento de Veículos (PRV), são: construção da rota; regra de transição; atualização do feromônio e critério de parada.

A implementação proposta neste trabalho consiste em estabelecer m rotas distintas, com m átomos, de maneira que cada rota possua pelo menos um nó do grafo e cada nó pertença a apenas uma das m rotas.

Para iniciar o algoritmo tem-se k conjuntos com m formigas cada, de tal forma que ao final do processo de construção das rotas sejam apresentados k conjuntos de rotas, cada qual com m rotas. O (ANEXO 1) apresenta um exemplo ilustrativo do algoritmo.

Por se tratar de um algoritmo de construção, cada uma das m rotas de cada um dos k conjuntos serão geradas, partindo de um nó inicial e incrementando a cada iteração deste processo um novo nó a cada rota, de maneira que cada nó só pertença a uma única rota do seu conjunto. O incremento de um novo nó à solução em construção dá-se pela seleção probabilística dos nós que ainda não fazem parte de nenhuma rota em construção. O algoritmo foi construído de forma que as formigas partam simultaneamente de pontos distintos e aleatórios da região. As formigas guardam "registros" das trilhas já percorridas por elas e por suas companheiras, a fim de que nenhum nó seja alocado em mais do que uma rota (lista tabu). Ao final deste processo construtivo, conforme dito anteriormente, cada um dos k conjuntos apresentará m rotas distintas, assim o algoritmo apresentará k soluções. São calculados os custos de cada uma das k soluções que é dado pela soma dos custos de cada uma das suas m rotas, finalizando assim a primeira iteração.

Para iniciar a próxima iteração deve-se atualizar a matriz de feromônios, conforme estabelecido na equação (3.62). Novamente, constrói-se k soluções e calculam-se seus custos. Lembrando que a cada iteração, todas as formigas partem simultaneamente de pontos distintos. Este procedimento é repetido até que um critério de parada seja atingido. O critério de parada pode ser estabelecido pelo número de iterações ou por tempo de processamento. Neste trabalho optou-se pelo número de iterações.

Ao atingir o critério de parada, o algoritmo retorna a melhor solução, o conjunto com m rotas que apresentou o menor custo total.

Ao término, o conjunto de formigas que apresentar a melhor solução, apresentará a melhor divisão geográfica, logo, melhores átomos.

A implementação inicialmente foi realizada com k conjuntos, todos iguais, pois todos os átomos tinham mesmo número de pontos. Posteriormente, foi realizada uma mudança no algoritmo a fim de mudar o número de pontos dentro de cada rota, com átomos com número de pontos diferentes. A mudança visa melhorar a divisão geográfica de acordo com os chamados emergenciais.

Na (FIGURA 3.9), há um exemplo ilustrativo com apenas um conjunto contendo quatro formigas, na pesquisa tem-se k conjuntos com m formigas cada. Inicialmente, as formigas partem simultaneamente de pontos aleatórios, conforme mostra a (FIGURA 3.10). Na sequência, cada formiga calcula, por meio das probabilidades, o próximo ponto a ser visitado. O cálculo acontece simultaneamente. Observar as (FIGURAS 3.11, 3.12 e 3.13). O local com maior valor da probabilidade é o local para onde as formigas seguirão. Elas memorizam os pontos por onde elas próprias passaram, bem como das suas companheiras, afinal, elas não podem passar por onde as outras formigas passaram e nem repetir pontos já visitados. Ao final, quando cada formiga concluir sua rota, ou seja, fechar o seu átomo, calculam-se os custos de cada um e atualiza-se a matriz de feromônio, conforme mostra a (FIGURA 3.14).

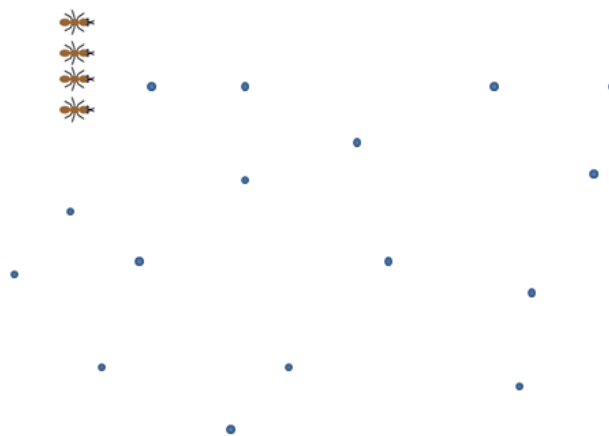


FIGURA 3.9 - UM CONJUNTO COM 4 FORMIGAS

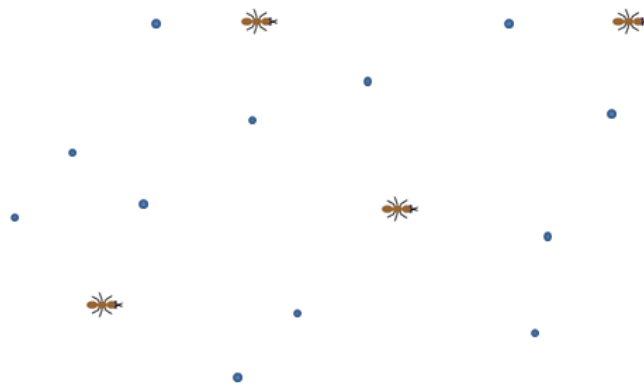


FIGURA 3.10 - FORMIGAS SAINDO SIMULTANEAMENTE E ALEATORIAMENTE

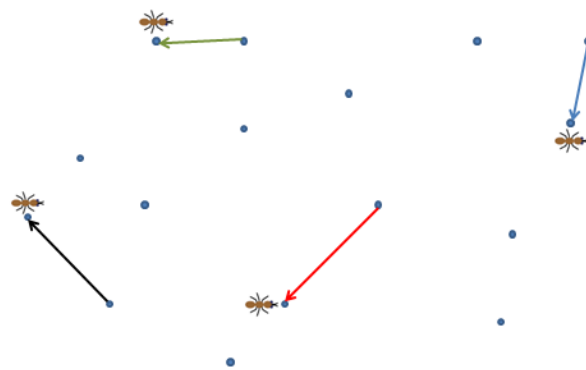


FIGURA 3.11 - CALCULANDO AS PROBABILIDADES

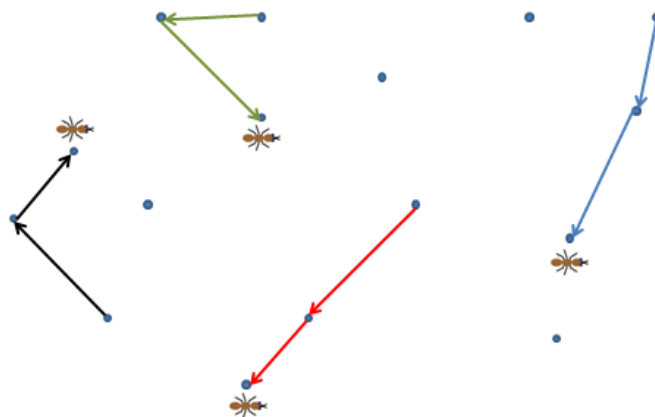


FIGURA 3.12 - CALCULANDO AS PROBABILIDADES

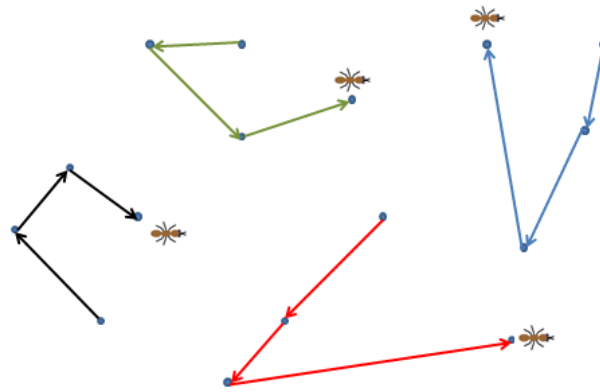


FIGURA 3.13 - CALCULANDO AS PROBABILIDADES

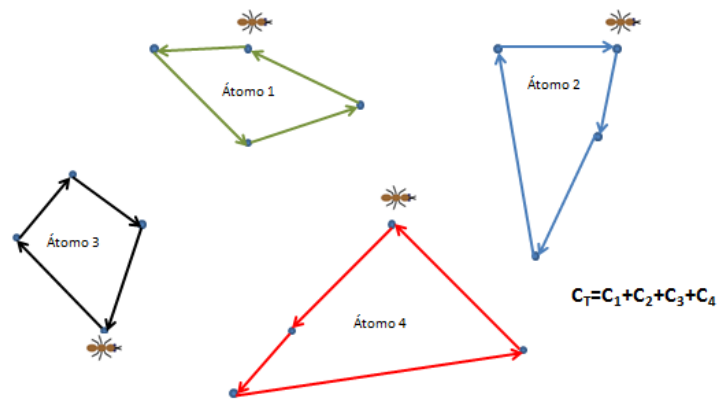


FIGURA 3.14 - FINAL DA PRIMEIRA ITERAÇÃO

A (FIGURA 3.14) mostra o final da primeira iteração, em que cada átomo possui o mesmo número de pontos. Calculam-se os custos e atualiza-se a matriz de feromônio.

A (FIGURA 3.15) apresenta um fluxograma do pseudo algoritmo baseado no ACO para múltiplos caixeiros viajantes. Tem-se, então, k conjuntos com m formigas cada. Todas as formigas saindo aleatoriamente e ao mesmo tempo, gerando dessa forma m átomos geográficos.

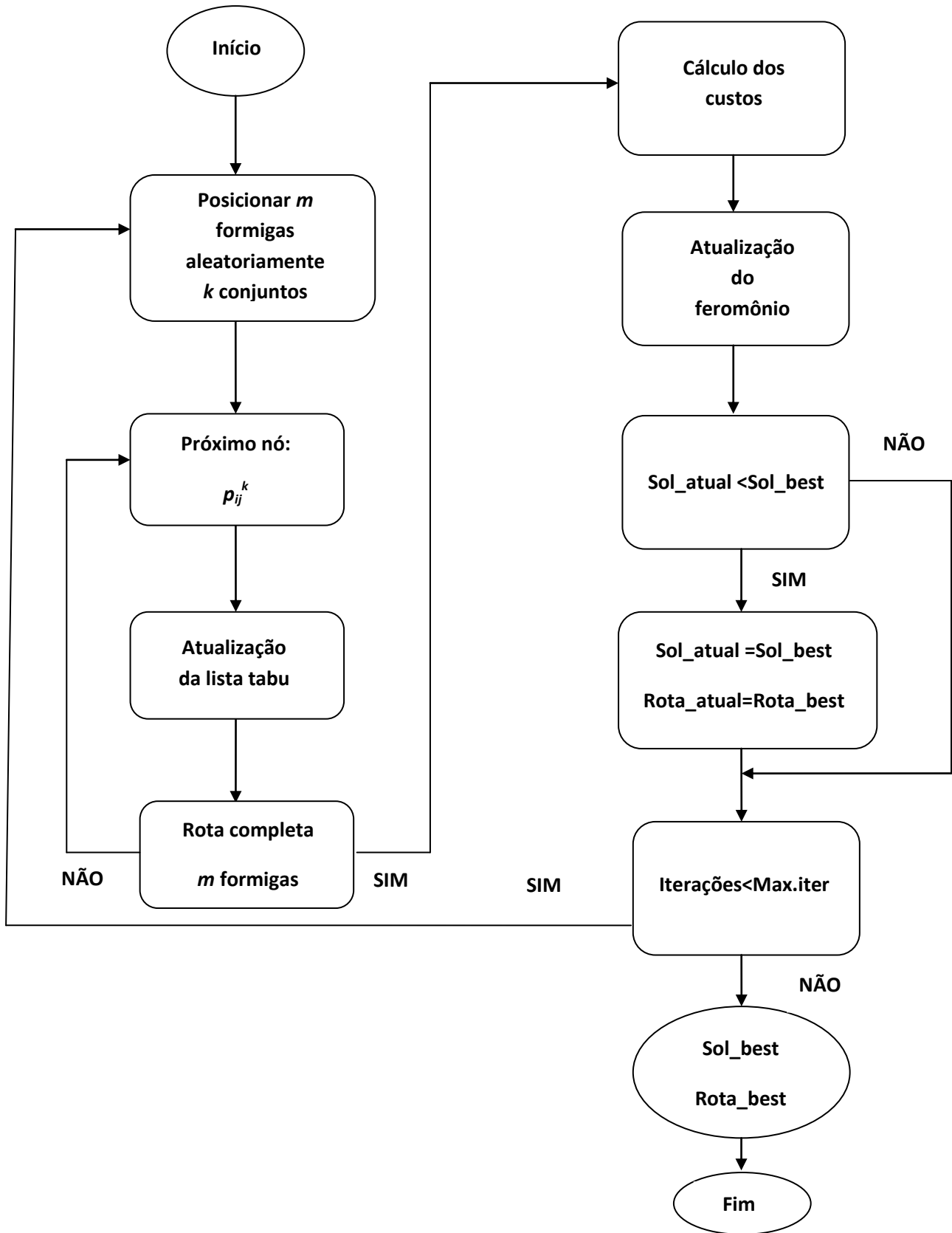


FIGURA 3.15 - FLUXOGRAMA REPRESENTANDO O PSEUDO ALGORITMO BASEADO NO ACO PARA MÚLTIPLOS CAIXEIROS VIAJANTES

4 MODELAGEM, VALIDAÇÃO E APLICAÇÃO DAS HIPÓTESES DO MODELO HIPERCUBO DE FILAS

Nesta seção é apresentada, por meio de um estudo de caso, a modelagem de um sistema de atendimento emergencial (SAE) móvel de filas, considerando suas características específicas de atendimento diferenciado. A primeira característica analisada trata do tipo de atendimento que vai desde casos graves que necessitam de veículos equipados com UTI, até simples remoções ou consultas. Outro aspecto relevante é a existência de fila de espera. Caso todos os servidores encontrem-se ocupados, o chamado entrará numa fila com um número pré-determinado de usuários. Vale lembrar que no SAE móvel a fila de espera possui capacidade infinita.

A coleta dos dados necessários para a validação e aplicação das nove hipóteses do modelo hipercubo de filas, ferramenta que irá modelar o SAE móvel tanto em Curitiba quanto em qualquer cidade do país com essas características, será feita também com estudo de caso.

Ainda nesta seção encontra-se a adaptação proposta ao algoritmo *Ant Colony Optimization* (ACO) com seus parâmetros e suas variações. O uso desta heurística modificada tem por finalidade comparar cenários distintos quanto à divisão dos átomos geográficos, obtendo dessa forma diferentes medidas de desempenho fornecidas pelo modelo hipercubo de filas, sempre visando reduzir o tempo de resposta.

Finalmente, foram determinados os possíveis estados de equilíbrio do sistema, bem como as respectivas equações de balanço para o SAE móvel.

4.1 SISTEMAS DE ATENDIMENTO EMERGENCIAL MÓVEL EM CURITIBA

Um sistema de atendimento emergencial de boa qualidade está diretamente relacionado à satisfação dos usuários do sistema, e consequentemente, a uma melhor qualidade de vida, seja ela individual ou coletiva (BELTRAMI, 1977; SWERSEY E INGALL, 1986). Sabe-se, no entanto, que para a obtenção de um bom atendimento, faz-se necessário que se planeje um conjunto de fatores, não apenas no aspecto financeiro, relacionado aos recursos, mas também no que diz respeito a um planejamento bem conduzido por parte do gerente do sistema, com tomadas de decisões acertadas. Segundo Richardson *et al.* (1995), para se planejar um sistema

de transporte é necessário diagnosticar o problema, identificar as principais causas, e, finalmente, compreender o seu comportamento pelas técnicas conhecidas.

Para avaliar um SAE móvel, é necessária, inicialmente, uma descrição verídica do funcionamento do sistema. Quanto mais precisas forem as informações, mais criteriosa será a análise do problema. Após esta análise, identifica-se a melhor ferramenta para aplicação, seguida da avaliação dos resultados obtidos com a técnica sugerida.

Um dos fatores críticos nesse tipo de pesquisa é a coleta de dados precisos, informações seguras e confiáveis.

Na cidade de Curitiba, além da empresa que forneceu os dados para execução deste trabalho, também prestam serviço de atendimento de emergência móvel o SAMU e o Corpo de Bombeiros. Primeiramente, buscou-se contato com o SAMU, a fim de aplicar o modelo proposto e analisar o cenário atual, porém, seus administradores não mostraram interesse na pesquisa.

Diante da recusa por parte dos administradores do SAMU Curitiba, buscou-se trabalhar com essa empresa, tendo em vista o interesse por parte do gerente em participar da pesquisa. A empresa forneceu as informações necessárias para a aplicação da metodologia proposta, atendendo, dessa forma, os objetivos propostos neste estudo.

4.2 DESCRIÇÕES DOS ATENDIMENTOS

A empresa conta com um sistema de informatização preciso e atualizado. Os dados desta pesquisa foram coletados, conforme mencionado na seção 2, nos anos de 2012 e 2013. Inicialmente, verificaram-se as taxas de chegada dos chamados ao longo do ano, considerando os meses, as semanas e os dias. Os tipos das chamadas, que são: urgência, emergência, remoção ou consulta e outros. Posteriormente, com a finalidade de validar as hipóteses do modelo hipercubo de filas, foram analisados os dados de sete dias da semana do mês de fevereiro de 2013. A escolha desse mês ocorreu de forma aleatória.

Dados estatísticos retirados das observações realizadas ao longo do ano (maio 2012 a abril de 2013), referente à quantidade de atendimentos, mostram que não há diferenças significativas entre os meses, exceto em meses de férias

escolares, conforme dados no gráfico da (FIGURA 2.6). O mesmo aconteceu dentro de cada mês, excetuando para os meses com feriados.

Com a finalidade de aplicar a metodologia proposta nesta tese para registrar as diferenças com relação à quantidade de chamados ao longo do dia, fez-se necessário determinar os períodos de picos dentro de cada dia analisado e o intervalo com maior número de chamados por unidade de tempo.

Pode-se observar que existem dois horários com maior número de chamados, um em torno das 10h e outro em torno das 18h, conforme a (FIGURA 2.7) da seção 2. Os gráficos dos chamados nos demais dias apontam horários de pico semelhantes, com desvios pouco significativos. Segundo dados estatísticos referentes aos anos de 2012 e 2013, fornecidos pela própria empresa, o comportamento dos chamados, bem como os tipos dos chamados possui uma constância, excetuando situações extraordinárias como feriados, férias ou epidemias, como no caso da gripe H1N1.

4.3 PERÍODOS DE PICO

O SAE móvel em Curitiba, de acordo com dados fornecidos pela empresa, demonstrou que o horário de maior congestionamento, maior número de chamadas por intervalo de tempo, ocorre nos períodos da manhã e tarde. O período da noite, compreendido entre 0 e 8 horas, apresenta o menor número de chamadas por intervalos de tempo. Neste estudo de caso, foi analisado o período da manhã, que é definido no período entre as 9 e 16 horas, e o período da tarde, definido das 17 às 23 horas. A escolha dos períodos se deu de forma a manter ou melhorar a principal característica do SAE móvel: a rapidez.

A rapidez no tempo de resposta é muito importante para o sistema, por isso, os indicadores de desempenho fornecidos pelo modelo hipercubo de filas devem apoiar as tomadas de decisão por parte do gerente do sistema em todos os períodos do dia, mas principalmente nos períodos críticos.

De acordo com os dados do (GRÁFICO 4.1) que representa as médias do horário de chegada de ocorrências por átomo, percebe-se que o horário com maior número de chamadas em todos os átomos, segundo a divisão usual (BPM), está compreendido entre as 13 e 14 horas. Contudo, faz-se necessário uma análise

consistente e efetiva para determinar de fato o intervalo que define o pico de chegada dos chamados.

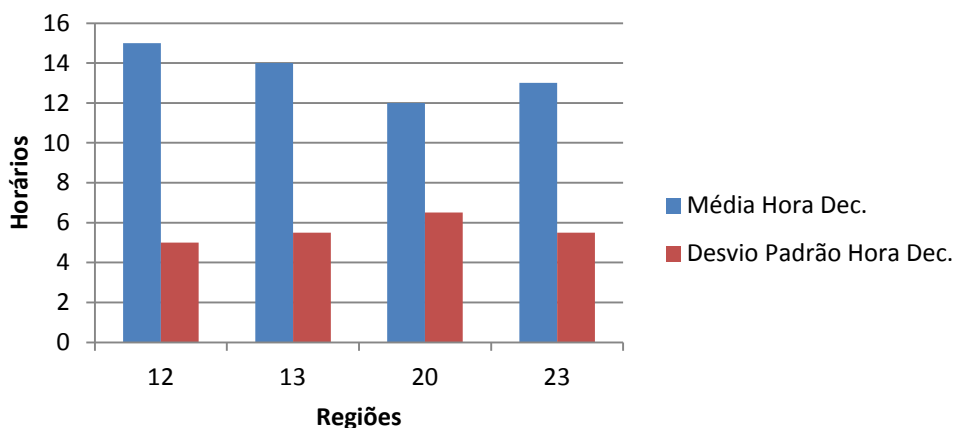


GRÁFICO 4.1 - MÉDIA DO HORÁRIO DE CHEGADA DOS CHAMADOS

Cabe ressaltar que, o período de pico foi obtido a partir do número de chamados obtidos dentro de um intervalo de sete dias. Foram analisados no intervalo das 9 às 16 horas, vários subintervalos.

O período de pico para a manhã ficou caracterizado pelo intervalo de tempo compreendido entre as 10 e as 14 horas, uma vez que este apresentou maior taxa de chegada dos chamados e consequentemente a maior média. Constataram-se em todos os outros subintervalos analisados que as taxas de chegada foram inferiores ao número de chamados no intervalo mencionado como pico.

O mesmo procedimento foi realizado para o período da tarde, no qual se constatou que o intervalo de pico ocorre entre as 18 e 21 horas, período considerado crítico do ponto de vista do sistema. Os mesmos veículos estavam à disposição tanto no período da manhã quanto no período da tarde, gerando dessa forma uma possível análise comparativa entre as medidas de desempenho obtidas pelo modelo nos dois períodos de pico.

4.4 VALIDAÇÃO DAS HIPÓTESES DO MODELO HIPERCUBO DE FILAS

Considerando as características do SAE móvel, faz-se necessária a verificação e validação das nove hipóteses do modelo hipercubo de filas, conforme descrita na seção 3, a fim de aplicá-lo ao problema em questão.

4.4.1 Divisão da área em N_A átomos geográficos

Existem diversas maneiras de dividir uma região, por exemplo, divisão por setores policiais, por bairros, divisão política, entre outras. Neste estudo, a divisão foi realizada de duas maneiras distintas a fim de compararmos as diferentes medidas de desempenho obtidas pelos cenários propostos. Primeiramente, quatro átomos geográficos foram definidos com a divisão considerada pelos Batalhões da Polícia Militar (BPM) de Curitiba. São eles: 12º BPM, 13º BPM, 20º BPM, 23º BPM.

Dentro de cada átomo houve uma divisão devido à particularidade dos tipos de atendimento do SAE móvel, que os classifica em: casos graves (emergência) e não tão graves (urgência). Os casos considerados leves, como as remoções e consultas, não foram considerados neste trabalho. Assim, com quatro átomos e dois tipos de atendimento, obtêm-se oito átomos geográficos.

A segunda divisão proposta aqui neste estudo foi obtida com a utilização da meta-heurística baseada no algoritmo *Ant Colony Optimization* (ACO), inicialmente utilizada para resolver PCV. Propôs-se uma adaptação no algoritmo ACO, a fim de aplicá-lo ao Problema de Múltiplos Caixeiros Viajantes (PMCV). A implementação consiste basicamente em estabelecer k conjuntos com m formigas, nos quais cada formiga criará rotas distintas.

Dessa forma, cada caixeiro ou formiga irá definir um conjunto de pontos do espaço geográfico (rotas), e cada um desses conjuntos de pontos ou cada rota, deverá definir um átomo geográfico. Assim, nesta nova divisão, de acordo com as prioridades de atendimentos, tem-se oito átomos geográficos. Nesse novo algoritmo as m formigas partem simultaneamente de pontos aleatórios do espaço geográfico, neste caso cada chamada representa um ponto do espaço geográfico.

O algoritmo trabalha com várias formigas em vários grupos de busca ao mesmo tempo. Incrementam-se a cada passo deste processo um novo ponto (nó) a cada rota. O acréscimo de um novo nó à solução, ocorre pela seleção probabilística dos pontos que ainda não fazem parte de nenhuma rota em construção. As formigas guardam na memória os pontos por onde já passaram bem como suas companheiras, uma vez que não é permitido alocar um nó em mais de uma rota.

Inicialmente, todas as formigas criaram rotas com o mesmo número de pontos. Posteriormente, com a nova implementação computacional desta tese, cada formiga de cada um dos conjuntos cria rotas de tamanhos distintos, com número de

pontos diferentes, mas sempre utilizando todos os pontos. Ao final desse processo construtivo, cada um dos k conjuntos apresenta m rotas distintas e k soluções. São calculados os custos de cada uma das k soluções, que é dado pela soma dos custos de cada uma das suas m rotas, finalizando assim a primeira iteração.

A próxima iteração do novo algoritmo é iniciada após a atualização da matriz de feromônios. Novamente, constrói-se k soluções e calculam-se seus custos. Lembrando que a cada iteração, todas as formigas partem simultaneamente de pontos aleatórios. Repete-se o procedimento até que um critério de parada seja atingido. Neste trabalho, foi estipulado um número máximo de 500 iterações, gerando ao final do processo, conjuntos de pontos (átomos) que são divisões geográficas diferentes das divisões já existentes. Vários testes foram realizados variando todos os parâmetros do algoritmo, os melhores resultados foram obtidos com $\beta = 2$, $\rho = 0,5$ e $\alpha = 1$.

Não se tem conhecimento da aplicação do algoritmo ACO juntamente com o modelo hipercubo de filas, estritamente descritivo, com a finalidade de criar átomos independentes. O objetivo de adaptar uma meta-heurística ao modelo é o de analisar outras medidas de desempenho fornecidas pelo modelo hipercubo e compará-las também ao cenário original.

4.4.2 Processo de chegada dos chamados

Foram analisados os processos de chegada dos chamados no período de pico, contemplando apenas os chamados de urgência e emergência. A amostra contém sete dias, com oito veículos em serviço. Têm-se, inicialmente dois cenários distintos, um relacionado à divisão dos átomos pelo BPM, e outro, à meta-heurística baseada no ACO.

A primeira análise é feita para o período da manhã (10 às 14 horas) com as duas divisões geográficas. A primeira divisão é a do BPM e a segunda divisão baseada no ACO, com rotas de tamanhos iguais e com rotas de tamanhos diferentes. Posteriormente, analisou-se o período da tarde (18 às 21 horas) com a divisão usual. A (TABELA 4.1) apresenta o Cenário 1, que mostra os resultados obtidos quanto ao número de chamados no período considerado, com a divisão segundo o BPM e no horário compreendido entre 10 e 14 h.

TABELA 4.1 - PROPORÇÃO DE CHAMADOS EM CADA ÁTOMO DO SISTEMA DO CENÁRIO 1

Átomos	Nº chamado	Proporção (p_{jk})
1 - 12A	20	0,096
2 - 12B	42	0,202
3 - 13A	17	0,082
4 - 13B	31	0,149
5 - 20A	26	0,125
6 - 20B	54	0,260
7 - 23A	6	0,029
8 - 23B	12	0,058
Total	208	1,000

Com o objetivo de verificar se o número de chegadas de chamadas em todos os átomos segue o padrão de Poisson, foi realizada uma análise em dados da amostra. Verificou-se que o número de chamadas segue o padrão de Poisson de contagem com incrementos independentes. Calculou-se a diferença entre o horário de chegada de cada chamado em cada átomo, o que gerou uma tabela para cada exemplo da amostra. Essa hipótese pode ser testada utilizando-se a propriedade de decomposição de processo de Poisson (WOLFF,1989), que diz: Seja $Y \sim P_0(\lambda)$ um processo de Poisson com parâmetro $\lambda > 0$ particionado aleatoriamente em k classes, onde $\sum p_j$ ($j = 1, 2, 3, \dots, k$) é a probabilidade de ocorrência da classe j , e $\sum Y_j$ é a frequência da classe j , com $p_j = 1$ e $Y_j = Y$. Então as variáveis aleatórias $Y_1, Y_2, Y_3, \dots, Y_k$ são mutuamente independentes, com $Y_j \sim P_0(\lambda \cdot p_j)$, $j = 1, 2, 3, \dots, k$.

Segundo Takeda (2000), esse resultado pode ser melhor visualizado de acordo com o esquema da (FIGURA 4.1):

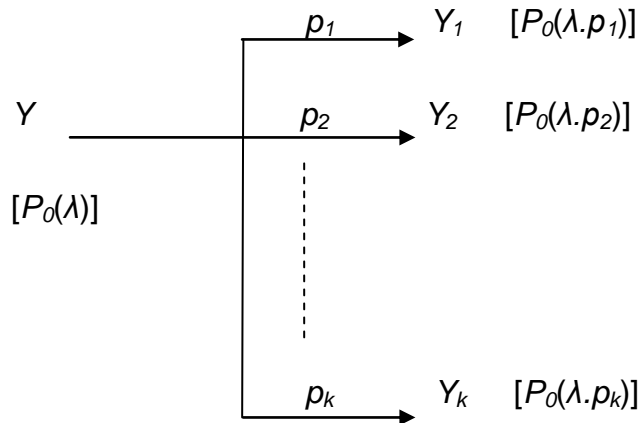


FIGURA 4.1 - DECOMPOSIÇÃO DE UM PROCESSO DE POISSON DE PARÂMETRO $\lambda > 0$

Com isso, testou-se a hipótese de que os intervalos médios entre as chamadas sucessivas apresenta uma distribuição exponencial, com parâmetro $(\lambda)^{-1}$. Os testes de aderência que foram realizados com a utilização do *software Arena* e o resultado é mostrado na (FIGURA 4.2) com o histograma da amostra. A partir dos resultados obtidos conclui-se que o número de chamados por intervalo de tempo, constitui um processo de Poisson, com parâmetro $\lambda > 0$ (ANEXO 2).

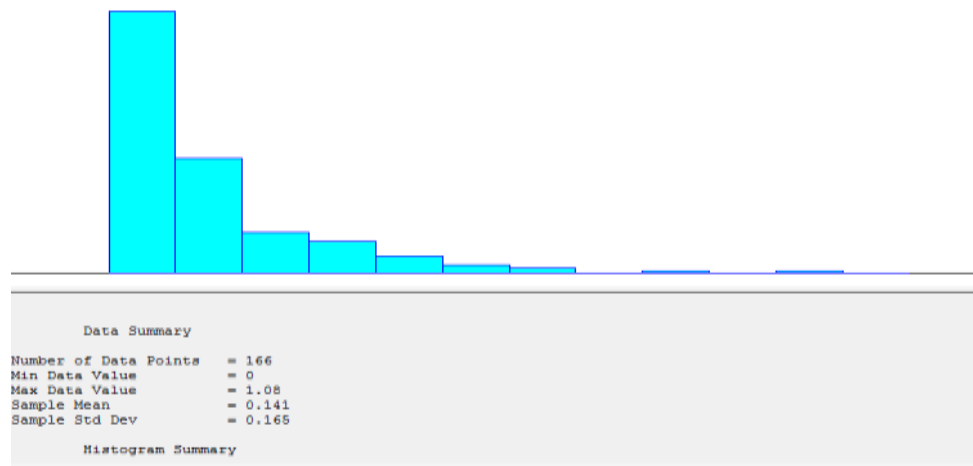


FIGURA 4.2 - HISTOGRAMA DA AMOSTRA DE SETE DIAS ANALISADOS

4.4.2.1 Cálculo dos λ_{jk}

Para o cálculo das taxas médias de chegadas dos chamados no sistema, foi considerado o período de pico, entre as 10 e 14 horas, e os intervalos médios de chegadas sucessivas para os dados da amostra. A (TABELA 4.2) apresenta o intervalo médio entre as chamadas, o tempo decorrido entre as chamadas sucessivas, o desvio padrão e o coeficiente de variação para exemplo da amostra.

TABELA 4.2 - INTERVALOS MÉDIOS ENTRE AS CHAMADAS SUCESSIVAS NO SISTEMA NO PERÍODO DE PICO

Dia	Intervalo médio (minutos)	Desvio-Padrão (minutos)	Coeficiente- Variação
1	6,05	5,31	1,33
2	10,5	10,79	1,66
3	10,81	12,46	1,78
4	6,81	5,69	1,03
5	9,72	8,6	1,43
6	6,53	5,94	1,19
7	8,44	7,56	1,26
Média	8,408	8,05	1,38

De acordo com os dados dispostos na tabela anterior, os desvios padrão são da ordem de grandeza das médias e os coeficientes de variação, na maioria dos dias, são próximos de um, com exceção do segundo e do terceiro dia, pelo fato de serem os dias com menores números de chamadas e conseqüentemente poucos dados. Isso só vem a reforçar a hipótese de que o intervalo de tempo entre as chegadas sucessivas dos chamados deve ser de fato exponencialmente distribuído, e que os chamados seguem uma distribuição de Poisson na chegada entre os chamados sucessivos (SOUZA, 2010). Pode-se verificar ainda na (TABELA 4.2) que a média do intervalo de tempo entre os chamados, em minutos, é de 8,4 e, conseqüentemente, a média por hora é de 0,14.

Foram realizados os testes Qui-Quadrado e Kolmogorov-Smirnov para verificar a aderência da amostra a uma distribuição exponencial, também no *software* Arena, conforme (FIGURA 4.3).

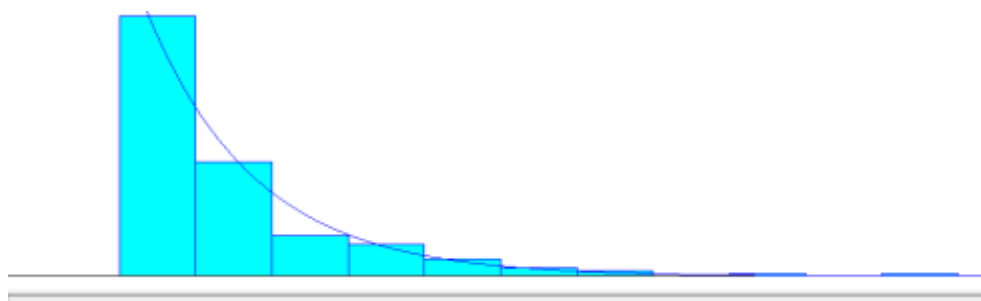


FIGURA 4.3 - HISTOGRAMA DA AMOSTRA E DISTRIBUIÇÃO

Concluiu-se, assim, que os intervalos de tempo entre as ocorrências consecutivas das chamadas seguem uma distribuição exponencial. Logo, o número

de vezes em que acontece uma chamada ao longo desse intervalo de tempo possui uma distribuição de Poisson. Mostrando que essa hipótese foi satisfeita.

Dessa forma, considerando que os chamados chegam independentemente em cada átomo e de acordo com o processo de Poisson, tem-se a segunda hipótese do modelo verificada. As taxas médias de chegada dos chamados podem ser calculadas por $\lambda_{jk} = \lambda \cdot \rho_{jk}$ ($j = 1, 2, 3, \dots, k$). A (TABELA 4.3) mostra esses valores calculados.

TABELA 4.3 - TAXAS MÉDIAS DE CHEGADA DOS CHAMADOS PARA CADA ÁTOMO POR PERÍODO

Átomos	Nº chamado	Proporção (ρ_{jk})	λ_{jk}
12A	20	0,09	0,69
12B	42	0,20	1,44
13A	17	0,08	0,58
13B	31	0,15	1,06
20A	26	0,13	0,89
20B	54	0,26	1,85
23A	6	0,03	0,21
23B	12	0,06	0,41
Total	208	1,00	7,14

4.4.3 Tempo de viagem

A obtenção dos tempos de viagem de cada veículo para cada átomo é de fundamental importância para a aplicação do modelo. Contudo, não é um dado de fácil acesso no sistema. Para a realização desta pesquisa, os dados foram fornecidos em diversas planilhas para que se calculassem os tempos de viagem de cada veículo. No entanto, podem surgir imprecisões, pois, os veículos após realizarem atendimentos, nem sempre retornam à base, podendo realizar outro atendimento em seguida. A (TABELA 4.4) mostra os tempos médios de viagem entre todos os átomos para o período de pico da manhã. Em problemas em que não é possível encontrar esses tempos, faz-se necessário o cálculo a partir de conceitos de probabilidade geométrica (LARSON e ODONI, 2007).

TABELA 4.4 - TEMPOS MÉDIO DE VIAGEM EM MINUTOS ENTRE SUBÁTOMOS OBTIDOS A PARTIR DE DADOS DA EMPRESA

Átomo	12A	12B	13A	13B	20A	20B	23A	23B	Média
12A	14,2	14,2	14,2	14,2	17,8	17,8	17,4	17,4	15,9
12B	14,2	14,2	14,2	14,2	17,8	17,8	17,4	17,4	15,9
13A	14,2	14,2	13,7	13,7	14,4	14,4	14	14	14,1
13B	14,2	14,2	13,7	13,7	14,4	14,4	14	14	14,1
20A	17,8	17,8	14,4	14,4	16,5	16,5	15,3	15,3	16
20B	17,8	17,8	14,4	14,4	16,5	16,5	15,3	15,3	16
23A	17,4	17,4	14	14	15,3	15,3	16	16	15,7
23B	17,4	17,4	14	14	15,3	15,3	16	16	15,7

Verifica-se, na (TABELA 4.4), que o tempo de deslocamento dos veículos dentro do próprio átomo é diferente de zero, isso acontece porque os átomos possuem dimensões consideráveis. Logo os veículos percorrem grandes distâncias mesmo quando realizam atendimento dentro do próprio átomo.

4.4.4 Servidores

A empresa conta com uma frota de oito veículos, dentre os quais quatro são utilizados para atendimento grave e quatro para atendimento não tão grave. O sistema opera com servidores espacialmente distribuídos, que podem se deslocar e atender a qualquer um dos átomos.

Todos os veículos possuem motoristas que também são socorristas, que podem participar do atendimento. Conforme já mencionado, os veículos basicamente diferem uns dos outros apenas em relação aos equipamentos que possuem.

4.4.5 Localização dos servidores

A localização dos servidores deve ser conhecida, ao menos probabilisticamente. Neste trabalho a matriz de localização dos servidores foi obtida junto ao sistema, de acordo com o critério utilizado pela empresa. Essa representação tornou-se bem simples, pois os veículos possuem bases definidas em seus átomos. A representação destas localizações dos veículos é feita de forma matricial $L=[l_{nj}]$, em que os elementos da matriz (l_{nj}) representam a probabilidade de um veículo estar localizado em determinado átomo no momento em que se

encontrar disponível. O sistema possui cinco bases de atendimento, e a localização dos veículos está representada na forma matricial, conforme (TABELA 4.5). A fim de facilitar a nomenclatura, os veículos 0, 1, 2 e 3 são os veículos que atendem prioritariamente casos graves (VG), e os veículos 4, 5, 6 e 7 atendem os casos não tão graves (VL).

TABELA 4.5 - MATRIZ DE LOCALIZAÇÃO DE VEÍCULOS

	12A	12B	13A	13B	20A	20B	23A	23B	Veículos
$L =$	0	0	0.5	0.5	0	0	0	0	0
	0	0	0	0	0.5	0.5	0	0	1
	0.5	0.5	0	0	0	0	0	0	2
	0.5	0.5	0	0	0	0	0	0	3
	0.5	0.5	0	0	0	0	0	0	4
	0.5	0.5	0	0	0	0	0	0	5
	0.5	0.5	0	0	0	0	0	0	6
	0.5	0.5	0	0	0	0	0	0	7

Como exemplo, de acordo com a tabela anterior, o veículo 0 encontra-se no átomo13, lembrando que cada átomo foi biparticionado, por isso, na matriz de localização a representação é dada por 0,5. De maneira análoga representa-se a localização dos demais veículos.

4.4.6 Despacho dos servidores

A política de despacho da empresa estabelece que apenas um servidor seja despachado para atender um chamado. Caso o chamado seja emergente, os veículos equipados com UTI são os primeiros a serem despachados para o atendimento, os demais veículos são *backups*. Caso o chamado seja urgente, será atendido preferencialmente por veículos mais simples, sendo os veículos equipados com UTI seus *backups*. No caso de todos os servidores encontrarem-se ocupados, o chamado entrará em uma fila de espera.

4.4.7 Política de despacho dos servidores

Há uma lista de preferências de despacho de servidores para cada átomo. Nesse caso em particular, como existem chamados graves e não tão graves, os

veículos preparados com UTIs, que são os veículos 0, 1, 2 e 3, são os primeiros a serem enviados num chamado grave (quando disponíveis), os demais veículos servem como *backup*. O mesmo procedimento serve para os outros veículos, são os primeiros a serem chamados para casos não tão graves e os veículos com UTIs são seus *backups*. A lista de preferência de despacho foi gerada de maneira aleatória, pelo método de geração de matrizes de despacho aleatórias (BURWELL *ET AL.* 1993), porém respeitando-se as particularidades do sistema. A (TABELA 4.6) apresenta um exemplo da matriz de preferência de despacho, na qual as linhas representam os átomos geradores dos chamados, e as colunas representam as preferências de despachos. O modelo hipercubo foi executado diversas vezes com várias matrizes. Ao final, realizaram-se os testes nos cenários distintos com a mesma matriz de preferência de despacho.

TABELA 4.6 - MATRIZ PREFERÊNCIA DE DESPACHO

Átomo	Preferência de despacho							
	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a	6 ^a	7 ^a	8 ^a
12A	2	3	1	0	4	5	6	7
12B	4	5	6	7	2	3	1	0
13A	0	1	3	2	5	6	7	4
13B	5	6	7	4	0	1	3	2
20A	1	2	0	3	6	7	4	5
20B	6	7	4	5	1	2	0	3
23A	3	0	2	1	7	4	5	6
23B	7	4	5	6	3	0	2	1

4.4.8 Tempo de atendimento

O tempo total de atendimento de um chamado é composto pelo somatório dos seguintes tempos: tempo de preparo do servidor (*setup time*), tempo de viagem do servidor até o local da ocorrência, tempo de execução do serviço junto ao usuário (tempo em cena) e o tempo de retorno à base. A (TABELA 4.7) apresenta os veículos que atendem casos graves VG e os veículos que atendem casos leves VL e os seus respectivos tempos médios de atendimento, coeficiente de variação, desvio padrão e as taxas médias de atendimento por hora dos veículos.

TABELA 4.7 - TEMPOS E TAXAS MÉDIAS PARA CADA VEÍCULO

Veículos	Tempo médio de atendimento (min)	Tempo médio de atendimento (hora)	Coefficiente de variação	Desvio padrão (min)	μ (hora)
0	63,98	1,07	0,36	20,43	0,94
1	56,69	0,94	0,59	32,61	1,06
2	48,28	0,80	0,76	26,02	1,24
3	39,17	0,65	0,84	23,53	1,53
4	31,20	0,52	0,64	25,60	1,94
5	41,30	0,68	0,98	31,31	1,46
6	42,50	0,71	1,01	31,51	1,41
7	45,00	0,75	0,33	13,42	1,33
V G	52,03	0,87	0,64	25,65	1,19
V L	39,88	0,66	0,74	20,46	1,54
Média	45,95	0,77	0,69	23,05	1,36

Verifica-se na (TABELA 4.7) que os desvios padrão e os tempos médios de atendimento são da mesma ordem de grandeza. Esse fato só vem confirmar que os intervalos de tempos de atendimento têm padrão exponencial, conforme indicam os testes feitos pelo do Método *Kolmogorov-Smirnov* para verificar a aderência da amostra a uma distribuição exponencial, realizados no *software Arena*.

As taxas médias μ_n mostradas na tabela anterior servem de base para a aplicação, desenvolvimento e obtenção dos resultados do modelo hipercubo de filas.

4.4.9 Dependência de atendimento em relação ao tempo de viagem

Com o objetivo de satisfazer a hipótese 9 do modelo, faz-se necessário verificar se os tempos médios de viagem são pequenos em relação ao tempo de serviço. Essa condição é normalmente atendida em serviços realizados nas áreas urbanas. Já nas áreas rurais, o tempo de deslocamento normalmente é relevante quando comparado ao tempo de atendimento.

Nas áreas rurais, justifica-se a necessidade de realizar um processo conhecido como calibração dos tempos médios de atendimento, ajustando os tempos médios de viagem para cada veículo, levando em consideração fatores geográficos que podem influenciar na viagem. Neste estudo, essa condição foi satisfeita conforme apresentado na (TABELA 4.8), que mostra a razão entre o tempo de viagem e o tempo de atendimento.

TABELA 4.8 - RELAÇÃO DO TEMPO DE VIAGEM COM O TEMPO DE ATENDIMENTO

Veículos	Tempo médio de atendimento (minutos)	Tempo médio de viagem (minutos)	Tempo viagem:Tempo atendimento
0	63,97	12,60	0,20
1	56,69	13,00	0,23
2	48,28	14,00	0,29
3	39,16	14,60	0,37
4	31,20	12,40	0,40
5	41,30	16,00	0,39
6	42,50	16,90	0,39
7	45,00	10,00	0,22
Média G	52,02	13,50	0,27
Média L	39,88	15,00	0,32

Verificando-se a razão entre os tempos de viagem e de atendimento, observa-se que o valor máximo é de 40% do tempo, o que não invalida a hipótese 9 do modelo.

Nem sempre os sistemas reais aderem a todas as hipóteses do modelo. Existem diversos fatores que podem influenciar negativamente. Entre eles, estão as imprecisões dos dados bem como a pequena quantidade dos mesmos. Contudo, segundo Larson e Odoni (1981), esses fatores de forma alguma inviabilizam a aplicação do modelo. Deve-se, porém, fazer uma análise profunda do sistema, verificar os pontos positivos do modelo e mesmo sem aderir a todas as hipóteses, deve-se verificar a aplicabilidade.

Neste estudo de caso, foram satisfeitas todas as hipóteses do modelo hipercubo, conforme descrito anteriormente. O próximo passo é determinar os possíveis estados a fim de encontrar as equações lineares de equilíbrio, bem como os valores das probabilidades de equilíbrio para todo o sistema. Em seguida, obter e analisar as medidas de desempenho.

4.5 TRANSIÇÕES ENTRE ESTADOS E EQUAÇÕES DE EQUILÍBRIO

O modelo hipercubo de filas baseia-se na resolução de um sistema de equações lineares. São essas equações que fornecem as probabilidades de equilíbrio dos possíveis estados do sistema e automaticamente as medidas de

desempenho. Contudo, para a aplicação do modelo hipercubo de filas, faz-se necessário um número reduzido de veículos, pois, conforme já mencionado anteriormente, com o aumento do número de servidores há um aumento exponencial de estados do modelo.

A aplicação do modelo hipercubo tem a finalidade de realizar uma análise descritiva do sistema sob o ponto de vista das medidas de desempenho, tanto do administrador como do usuário do sistema.

Conforme descrito anteriormente, nesta pesquisa trabalhou-se com oito átomos e com oito veículos, o que nos fornecerá um número de equações lineares igual a 2^8 , ou seja, 256 equações e 256 incógnitas. Tem-se, ainda, a fila de espera limitada a um usuário por veículo, com oito equações, uma equação para cada estado, cujas incógnitas também são suas probabilidades de equilíbrio do sistema. Esse sistema é indeterminado, e uma maneira de eliminar esta indeterminação é introduzindo uma equação de normalização, ou seja, considerando que a soma das probabilidades de todos os possíveis estados do sistema deve ser igual a um. A equação 4.1 a seguir expressa esta relação:

$$p_{00...0} + p_{00...1} + p_{0...10} + \dots + p_{1...11} = 1 - p_Q \quad (4.1)$$

Onde:

p_Q é probabilidade de haver fila de espera.

A fim de obter as equações de equilíbrio deve-se considerar a matriz de preferência de despachos, gerada aleatoriamente. O sistema de equações de fila segue o mesmo padrão no sistema e não existe limite para a fila. Para facilitar a aplicação do modelo, a fila nesse estudo foi considerada finita, o número de usuários que poderão aguardar na fila é igual ao número de veículos, contudo, a empresa trabalha com fila ilimitada. Essa adaptação na fila não interfere significativamente nos resultados.

A empresa realiza, além dos serviços de urgência e emergência, serviços de remoções e consultas. São serviços considerados de baixa prioridade e são realizados desde que o sistema esteja disponível. O número de atendimento é bastante elevado, gerando uma grande sobrecarga no sistema e esse atendimento na maioria das vezes não será interrompido na chegada de uma chamada de alta complexidade.

Nesta pesquisa não foram considerados os casos de baixa prioridade, ou seja, remoções e consultas. Foram tratados apenas os dados relacionados à urgência e emergência.

Quanto à chegada dos chamados no sistema, a probabilidade de que cheguem dois chamados simultaneamente é considerada nula. Também é considerada nula a probabilidade de dois servidores ficarem livres simultaneamente. Na prática, não se despacha mais de um veículo para atender a um chamado. Com isso, em qualquer transição é permitida apenas um passo por vez.

Com a finalidade de ilustrar o processo, a (FIGURA 4.4) mostra um exemplo do diagrama de fluxo do estado (00000000), em que o sistema encontra-se com os oito veículos desocupados, passando para os demais possíveis estados. Deve-se considerar a preferência de despacho da matriz, pois os veículos são despachados na ordem em que estão listados e por sua vez depende das suas respectivas taxas de transição.

De maneira análoga construiu-se todos os demais possíveis estados. As respectivas probabilidades de equilíbrio, além da fila de espera. As taxas de transições de um estado para outro são ascendentes ou descendentes.

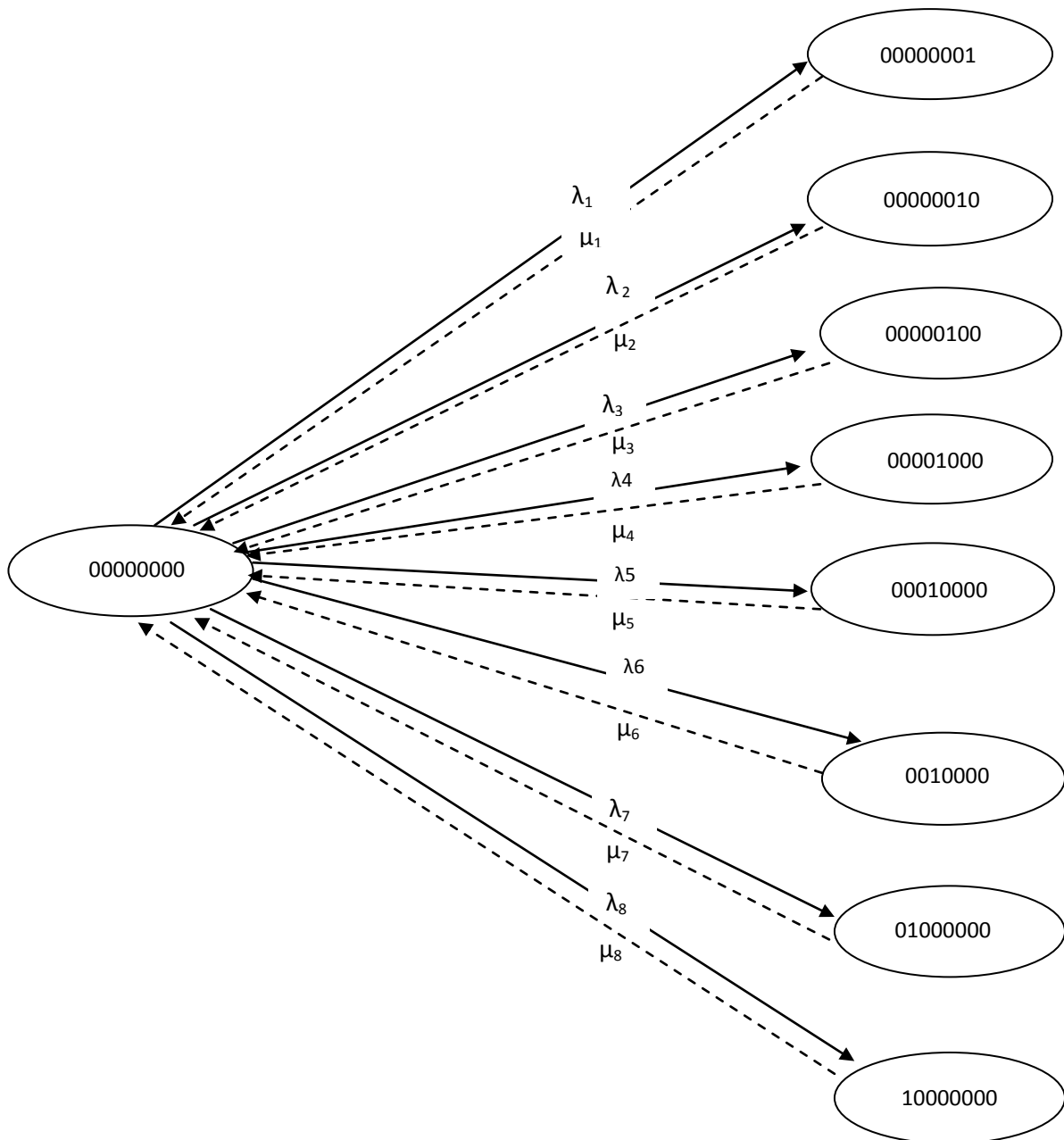


FIGURA 4.4 - DIAGRAMA DE FLUXO PARA O ESTADO VAZIO (00000000)

A solução do modelo é dada pela construção das equações de equilíbrio do sistema, que são definidas supondo-se que o sistema atinja o equilíbrio, pois o fluxo que entra neste estado deve ser igual ao fluxo que sai dele. Em um sistema não saturado, com capacidade de fila infinita, as probabilidades de estado do modelo hipercubo são calculadas a partir das equações de balanço, construídas a partir dos possíveis estados.

A fim de ilustrar o cálculo das equações de equilíbrio, analisa-se apenas a situação inicial em que todos os servidores encontram-se desocupados, logo, os possíveis estados ocorrerão se um dos veículos passar a ocupado, gerando nesse caso oito possíveis estados, tendo visto que apenas um passo poderá ocorrer.

A equação de balanço para o estado vazio (00000000), representado na (FIGURA 4.4) está descrita na equação 4.2 a seguir e o mesmo raciocínio foi utilizado para a construção das demais equações.

$$\begin{aligned} & [(\lambda_1) + (\lambda_2) + (\lambda_3) + (\lambda_4) + (\lambda_5) + (\lambda_6) + (\lambda_7) + (\lambda_8)] \cdot p_{00000000} = \\ & \mu_1 \cdot p_{00000001} + \mu_2 \cdot p_{00000010} + \mu_3 \cdot p_{00000100} + \mu_4 \cdot p_{000001000} + \\ & \mu_5 \cdot p_{00010000} + \mu_6 \cdot p_{00100000} + \mu_7 \cdot p_{01000000} + \mu_8 \cdot p_{10000000} \end{aligned} \quad (4.2)$$

Além das equações de balanço, deve-se considerar ainda as equações dos estados referentes à fila de espera no sistema. Dessa forma, têm-se todas as probabilidades necessárias para a obtenção dos indicadores de desempenho, os quais serão analisados e comparados nos diversos cenários propostos neste estudo.

5 IMPLEMENTAÇÃO E RESULTADOS DO MODELO HIPERCUBO DE FILAS COM DIVISÃO GEOGRÁFICA USUAL E COM DIVISÃO BASEADA NO ACO

O modelo hipercubo de filas e a meta-heurística baseada no ACO foram implementados computacionalmente em linguagem *Visual Basic* e executados em um computador com processador Intel Core I5-2400, de 64 Bits, CPU 3.10GHz, memória RAM 8GB e sistema operacional *Windows 7*.

Neste trabalho, a aplicação do modelo hipercubo de filas teve por objetivo, a análise descritiva do SAE móvel a partir das medidas de desempenho fornecidas pelo modelo. Para tanto, foram analisados dois períodos críticos ao longo do dia, um no turno da manhã e outro no turno da tarde. O modelo hipercubo foi aplicado em duas divisões geográficas. A primeira divisão, realizada pelo BPM, recebeu a denominação de cenário 1; a segunda divisão – de fundamental importância para este trabalho por caracterizar a inovação da pesquisa obtida a partir da meta-heurística baseada no ACO – cenário 2. Para o cenário 2, foi feita uma subdivisão: inicialmente trabalhou-se com o mesmo número de pontos em cada átomo; posteriormente, foram definidos números diferentes.

5.1 CENÁRIO 1: PERÍODO DA MANHÃ, DAS 10 ÀS 14 HORAS

Primeiramente serão apresentados os resultados do cenário 1, período a manhã, considerando a divisão do BPM, conforme mostra a (FIGURA 5.1), no intervalo das 10 às 14 horas. O modelo forneceu a probabilidade de encontrar o sistema vazio ($p_{00000000}$) igual a 0,00507. A probabilidade de encontrar o sistema saturado ($p_{11111111}$), com usuários aguardando para serem atendidos, foi de aproximadamente 0,1328. E a probabilidade de pelo menos um servidor encontrar-se ocupado é de 0,8671.

o tempo médio de resposta dos veículos (preparação e viagem) obtidos na amostra coletada são apresentados na (TABELA 5.1).

TABELA 5.1 - *WORKLOAD* DOS VEÍCULOS E OS TEMPOS DE RESPOSTA

Veículos	Workload	Modelo (minutos)	Amostra (minutos)	Diferença minutos	Diferença %	Desvio Padrão
0	0,42	14,2	12,6	1,6	12,70	1,13
1	0,51	16,1	14,1	2	14,18	1,41
2	0,49	15,4	14	1,4	10,00	0,99
3	0,31	15,6	14,6	1	6,85	0,71
4	0,6	15,5	15	0,5	3,33	0,35
5	0,65	15,1	16	-0,9	-5,63	0,64
6	0,74	16,4	19	-2,6	-13,68	1,84
7	0,68	16,4	16	0,4	2,50	0,28
VSG	0,43	15,33	13,83	1,50	10,85	1,06
VSL	0,67	15,85	16,50	-0,65	-3,94	0,46
Total	0,55	15,59	15,16	0,42	2,80	0,30

A (TABELA 5.1) apresenta as diferenças em minutos, em porcentagem e também o desvio padrão. Pode-se observar que a diferença mais alta foi a do veículo 6, com 13,68%. Os demais resultados apresentaram-se bem próximos do resultado do sistema, o que valida a aplicação do modelo para o serviço. Quanto à carga de trabalho dos veículos, os resultados obtidos a partir do modelo confirmam os dados fornecidos inicialmente pelo sistema: os veículos que servem preferencialmente casos graves ficam menos tempo ocupados, em torno de 43%, enquanto os veículos que atendem a chamados leves tendem a ficar mais tempo ocupados, no caso 67%, conforme apresentado na (FIGURA 5.2), o que fornece uma média global de 55%, compatível com o sistema.

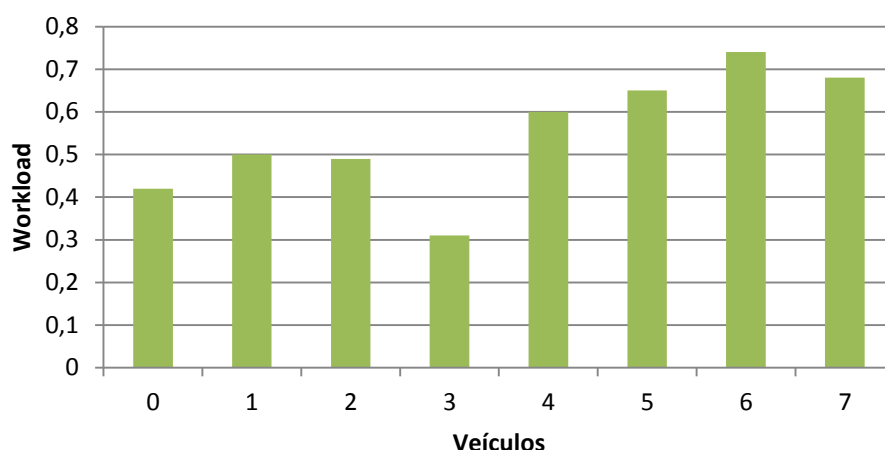


FIGURA 5.2 - WORKLOAD DOS VEÍCULOS

Pode-se também analisar, conforme apresentado na (TABELA 5.2), o tempo de resposta dos veículos, ou seja, o tempo de resposta para cada átomo em função da região onde ocorrem os chamados. Essa tabela, apresenta a diferença em minutos, em porcentagem e o desvio padrão entre o tempo fornecido pelo modelo e o tempo real (coletado na amostra).

TABELA 5.2 - TEMPOS MÉDIO DE RESPOSTA DE CADA ÁTOMO

Átomos	Modelo	Amostra	Diferença (min)	Diferença %	Desvio Padrão
1-12A	14,60	14,80	-0,20	-1,35	0,14
2-12B	14,50	16,40	-1,90	-11,59	1,34
3-13A	13,90	12,90	1,00	7,75	0,71
4-13B	14,20	16,80	-2,60	-15,48	1,84
5-20A	16,60	14,80	1,80	12,16	1,27
6-20B	17,40	19,20	-1,80	-9,38	1,27
7-23A	16,70	13,60	3,10	22,79	2,19
8-23B	17,10	20,10	-3,00	-14,93	2,12
VA	15,50	14,00	1,50	10,71	1,06
VB	15,80	18,1	-2,30	-12,71	1,63
Total	15,60	16,05	-0,45	-2,80	0,32

Verifica-se pelos dados contidos na (TABELA 5.2) que os tempos fornecidos pelo modelo e pela amostra apresentam desvios pouco significativos, exceto no átomo 23. Isso se deve ao fato de ser o átomo com o menor número de chamados (apenas 6 chamados graves). As diversas regiões geradoras dos chamados apresentam tempos médios de respostas próximas à do modelo, tanto em casos

graves como em casos leves. Os casos graves apresentam desvio padrão médio da ordem de “1”, enquanto os casos leves apresentam desvios da ordem de 1,6. A (FIGURA 5.3) mostra a comparação, modelo e amostra, para a distribuição dos tempos de respostas para cada um dos átomos.

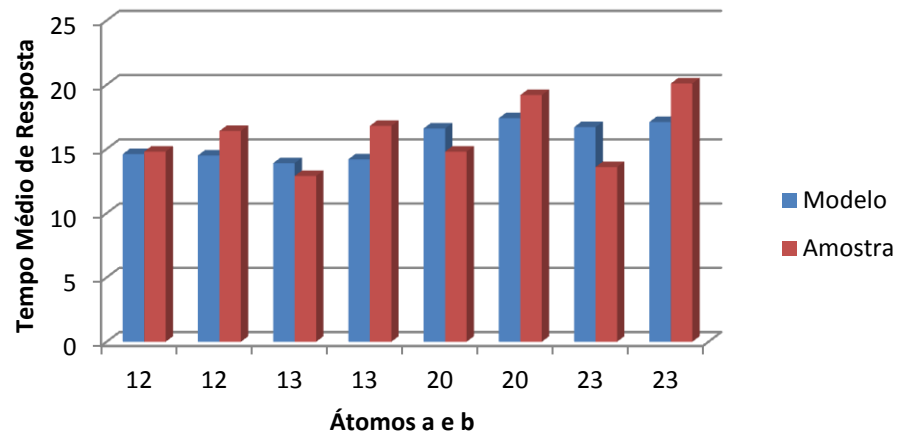


FIGURA 5.3 - TEMPOS MÉDIO DE RESPOSTA PARA CADA ÁTOMO

- b. A frequência de despacho f_{ij} que é a fração de todos os despachos do servidor i ao átomo j e pode ser decomposta em duas partes: $f_{ij}^{[1]}$, que corresponde à fração de despachos de um servidor i para um átomo j que não implica tempo de espera em fila para o usuário; e $f_{ij}^{[2]}$, que corresponde à fração de despachos de um servidor i para um átomo j sujeito à espera em fila.

A frequência de despachos fornecida através do modelo para o servidor i deslocar-se até o átomo j , é de 0,58 para a fração dentre todos os despachos que são interáreas de cobertura.

A fração dos despachos que são interáreas (*Backup*) de cobertura de cada um dos veículos encontra-se na (TABELA 5.3).

TABELA 5.3 - FRAÇÃO DOS DESPACHOS DE CADA VEÍCULO QUE SÃO INTERÁREAS DE COBERTURA

Veículos	<i>Backup</i>
0	0,401
1	0,493
2	0,478
3	0,297
4	0,573
5	0,622
6	0,720
7	0,359
Média VSG	0,417
Média VSL	0,568

Pode-se observar, conforme apresentado na (TABELA 5.3), que os veículos destinados ao atendimento de casos graves (VSG) são menos requisitados para atender a outros chamados (*backup*) do que os veículos que atendem preferencialmente casos leves (VSL). Os resultados obtidos através do modelo comprovam o que realmente acontece no SAE móvel.

A frequência de despacho de cada veículo, ou seja, a frequência de atendimentos em sua área de cobertura primária e a frequência de atendimentos *backup* podem ser observadas na (TABELA 5.4).

TABELA 5.4 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO

Veículos	Cobertura primária	<i>Backup</i>
0	0,533	0,401
1	0,478	0,493
2	0,574	0,478
3	0,785	0,297
4	0,520	0,573
5	0,577	0,622
6	0,390	0,720
7	0,548	0,359
VSG	0,592	0,417
VSL	0,508	0,568

De acordo com os resultados obtidos através do modelo hipercubo, conforme mostra a (TABELA 5.4), os veículos têm uma frequência de despacho bem

próxima quando se trata de atendimento primário. A diferença acontece apenas quando são despachados em segunda preferência (*backup*).

- c. Os tempos médios de viagem são obtidos através da matriz dos tempos de viagem no próprio sistema; podem, porém, sofrer influência de diversos fatores no dia a dia. Portanto, o tempo de deslocamento de um átomo i até um átomo j poderá não ser igual ao do caminho inverso.

O tempo médio global de viagem é uma importante medida de desempenho do ponto de vista do usuário do sistema.

O modelo forneceu um tempo médio de deslocamento para chamadas na fila de espera, em minutos, de 15,61. O tempo médio global de resposta do deslocamento, independentemente do átomo gerador do chamado e da equipe que irá atender ao chamado é de 15,64 minutos. Na empresa, esse tempo foi calculado em aproximadamente 15,67 minutos, uma diferença que permite a conclusão de que o modelo hipercubo é uma boa aproximação para o SAE móvel.

O tempo médio de deslocamento de cada veículo ao átomo j é outro importante indicador, pois reflete o nível de serviço oferecido pelo sistema.

A (TABELA 5.5) apresenta os tempos médios de viagem para cada átomo do problema abordado no modelo hipercubo e na amostra coletada, considerando tempos iguais de viagem dentro de cada átomo para a amostra.

TABELA 5.5 - TEMPOS MÉDIO DE VIAGEM PARA CADA ÁTOMO

Átomos	Modelo	Amostra	Dif (minutos)	Dif %	Desvio Padrão
1-12a	13,93	15,90	-1,97	-12,38	1,39
2-12b	16,50	15,90	0,60	3,77	0,42
3-13a	14,63	14,08	0,56	3,94	0,39
4-13b	16,72	15,08	1,65	10,91	1,16
5-20a	14,50	16,00	-1,50	-9,38	1,06
6-20b	14,10	16,00	-1,90	-11,88	1,34
7-23a	17,40	15,68	1,73	11,00	1,22
8-23b	17,06	15,68	1,39	8,84	0,98
Va	15,11	15,66	-0,55	-3,51	0,38
Vb	16,09	15,66	0,43	2,75	0,30
Total	15,60	15,66	-0,06	-0,38	0,04

Nessa tabela, a nomenclatura Va representa os veículos que atendem casos graves (átomo a) e Vb veículos que atendem casos básicos (átomo b). O modelo hipercubo de filas constitui uma ferramenta adequada ao problema, uma vez que as diferenças apresentadas foram pequenas, levando em consideração a imprecisão de alguns dados.

5.2 CENÁRIO 1: PERÍODO DA TARDE, DAS 18 ÀS 21 HORAS

Neste cenário, a probabilidade de encontrar o sistema vazio ($p_{00000000}$) é de 0,0056. A probabilidade de encontrar o sistema saturado, com usuários aguardando para serem atendidos, é de aproximadamente 0,120, enquanto a probabilidade de pelo menos um servidor encontrar-se ocupado é de 0,879.

Os resultados obtidos através do modelo para a carga de trabalho dos veículos (*Workload*), o tempo médio de resposta dos veículos e o tempo médio de resposta dos veículos obtidos na amostra encontram-se na (TABELA 5.6).

TABELA 5.6 - *WORKLOAD* DOS VEÍCULOS E OS TEMPOS DE RESPOSTA

Veículos	<i>Workload</i>	Modelo	Amostra	Diferença Minutos	Diferença %	Desvio Padrão
0	0,38	11,09	11,08	0,01	0,09	0,01
1	0,44	12,14	12,90	-0,76	-5,89	0,54
2	0,42	11,32	12,50	-1,18	-9,44	0,83
3	0,29	11,11	11,50	-0,39	-3,39	0,28
4	0,64	11,70	10,70	1,00	9,35	0,71
5	0,59	11,76	9,30	2,46	26,45	1,74
6	0,72	13,21	13,00	0,21	1,62	0,15
7	0,69	12,82	13,70	-0,88	-6,42	0,62
VSG	0,38	11,42	12,00	-0,58	-4,84	0,41
VSL	0,66	12,37	11,64	0,73	6,29	0,52
Total	0,52	11,89	11,82	0,08	0,65	0,05

Essa tabela apresenta as diferenças dos tempos de resposta, em minutos, em porcentagem e também o desvio padrão. Pode-se observar que o desvio mais alto foi o do veículo 5, com 26,45%. No entanto, os demais resultados apresentaram-se bem próximos do resultado do sistema, o que valida a aplicação do modelo de maneira satisfatória. Quanto à carga de trabalho dos veículos, é semelhante à dos veículos do período da manhã. Os veículos que servem

preferencialmente casos graves ficam menos tempo ocupados, em torno de 38%, enquanto os veículos que atendem a chamados leves tendem a ficar mais tempo atendendo, no caso 66%, o que nos forneceu uma média global de 52%, pouco inferior à apresentada no período da manhã, contudo próxima à do sistema, que é aproximadamente igual a 50%.

A (TABELA 5.7) apresenta o tempo de resposta para cada átomo obtido a partir do modelo e da amostra coletada, a diferença em minutos, em porcentagem e os respectivos desvios padrões.

TABELA 5.7 - TEMPOS MÉDIO DE RESPOSTA DE CADA ÁTOMO

Átomos	Modelo	Amostra	Diferença Minutos	Diferença %	Desvio Padrão
1-12a	10,70	11,30	-0,60	-5,31	0,42
2-12b	10,50	9,90	0,60	6,06	0,42
3-13a	10,90	11,50	-0,60	-5,22	0,42
4-13b	11,60	10,78	0,82	7,61	0,58
5-20a	12,70	13,25	-0,55	-4,15	0,39
6-20b	14,20	13,88	0,32	2,34	0,23
7-23a	11,70	11,00	0,70	6,36	0,49
8-23b	10,76	13,00	-2,24	-17,23	1,58
VSG	11,50	11,76	-0,26	-2,21	0,18
VSL	11,76	11,88	-0,12	-1,01	0,08
Total	11,63	11,82	-0,19	-1,61	0,13

Verifica-se, pela (TABELA 5.7), que os tempos fornecidos pelo modelo e pela amostra apresentam diferenças pouco significativas, exceto no átomo 23, mesmo problema apresentado no período da manhã.

A fração dos despachos que são interáreas (*Backup*) de cobertura de cada um dos veículos pode ser observada na (TABELA 5.8).

TABELA 5.8 - FRAÇÃO DOS DESPACHOS DE CADA VEÍCULO QUE SÃO INTERÁREAS DE COBERTURA

Veículos	<i>Backup</i>
0	0,367
1	0,425
2	0,406
3	0,275
4	0,623
5	0,580
6	0,704
7	0,680
VSG	0,368
VSL	0,646

Os veículos destinados ao atendimento de casos graves, analogamente ao que acontece no período da manhã, são menos requisitados para atender a outros chamados (*backup*) do que os veículos que atendem preferencialmente casos leves.

As frequências de despacho de cada veículo, em sua área de cobertura primária e a frequência de atendimentos *backup* podem ser observadas na (TABELA 5.9).

TABELA 5.9 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO

Veículos	Cobertura primária	<i>Backup</i>
0	0,429	0,367
1	0,596	0,425
2	0,502	0,406
3	0,933	0,275
4	0,452	0,623
5	0,643	0,580
6	0,405	0,704
7	0,887	0,680
VSG	0,615	0,368
VSL	0,596	0,646

A (TABELA 5.9) indica que, em se tratando de cobertura primária, existe um equilíbrio entre os veículos que atendem a chamados graves e os veículos que atendem a chamados leves. Por outro lado, quando atuam como *Backup*, os

veículos que atendem prioritariamente a chamados leves são mais requisitados do que os veículos que atendem a chamados graves.

O modelo forneceu um tempo médio de deslocamentos de 12,42 minutos para chamadas na fila de espera e um tempo médio global de resposta do deslocamento, independentemente do átomo gerador e da equipe que atendeu o chamado, de 11,94 minutos. Na empresa, esse tempo foi calculado em aproximadamente 12,18 minutos, uma diferença pequena, em torno de 2%, validando também nesse horário o modelo hipercubo de filas.

A (TABELA 5.10) apresenta os tempos médios de viagem aos átomos do problema abordado no modelo e na amostra sem levar em consideração as diferenças entre os tempos de viagem dentro do mesmo átomo.

TABELA 5.10 - TEMPOS MÉDIO DE VIAGEM DE CADA VEÍCULO PARA CADA ÁTOMO

Átomos	Modelo	Amostra	Diferença minutos	Diferença %	Desvio Padrão
1-12a	10,89	11,00	-0,11	-1,00	0,08
2-12b	12,72	11,00	1,72	15,64	1,22
3-13a	10,70	11,35	-0,65	-5,73	0,46
4-13b	11,74	11,35	0,39	3,44	0,28
5-20a	10,50	13,40	-2,90	-21,64	2,05
6-20b	11,62	13,40	-1,78	-13,28	1,26
7-23a	14,17	13,00	1,17	9,00	0,83
8-23b	10,76	13,00	-2,24	-17,23	1,58
Va	11,56	12,18	-0,62	-5,09	0,44
Vb	11,71	12,18	-0,47	-3,86	0,33
Total	11,64	12,18	-0,54	-4,45	0,38

Os resultados obtidos a partir do modelo também foram bem próximos dos resultados coletados na amostra. Apresentaram uma diferença percentual média de 4,45%.

5.3 COMPARAÇÃO DAS MEDIDAS DE DESEMPENHO DO CENÁRIO 1, PERÍODOS DA MANHÃ E DA TARDE

A (TABELA 5.11) mostra algumas medidas de desempenho nos dois horários de pico, das 10 às 14 horas (manhã) e das 18 às 21 horas (tarde).

TABELA 5.11 - COMPARAÇÃO *WORKLOAD* NOS DOIS PERÍODOS

10 às 14 horas			18 às 21 horas		Diferenças	
Veículos	<i>Workload</i> (manhã)	Tempo de resposta (minutos)	<i>Workload</i> (tarde)	Tempo de resposta (minutos)	minutos	%
0	0,42	14,2	0,38	11,09	3,11	28,04
1	0,51	16,1	0,44	12,14	3,96	32,62
2	0,49	15,4	0,42	11,32	4,08	36,04
3	0,31	15,6	0,29	11,11	4,49	40,41
4	0,6	15,5	0,64	11,70	3,8	32,48
5	0,65	15,1	0,59	11,76	3,34	28,40
6	0,74	16,4	0,72	13,21	3,19	24,15
7	0,68	16,4	0,69	12,82	3,58	27,93
VSG	0,43	15,33	0,38	11,42	3,91	34,24
VSL	0,67	15,85	0,66	12,37	3,48	28,13
Total	0,55	15,59	0,52	11,89	3,7	31,12

A partir da análise das medidas de desempenho, pode-se verificar que os *workloads* do período da manhã são maiores que os *workloads* do período da tarde, conforme mostra a (FIGURA 5.4), apesar de as taxas de chegada dos chamados serem praticamente iguais. Constata-se também que o tempo de resposta no período da tarde é em média 31,1% menor quando comparado ao do período da manhã. Isso se deve sobretudo ao fato de o trânsito ser menos intenso neste horário. Dessa forma, o período da tarde, apesar de, proporcionalmente, ser o período com maior número de chamadas, é o período em que os veículos atendem aos chamados mais rapidamente.

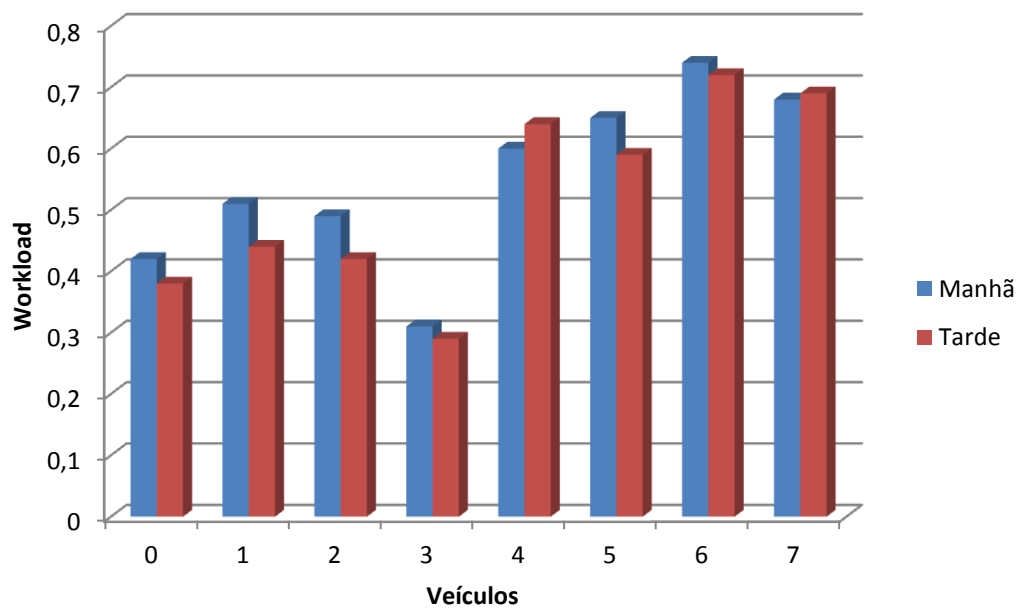


FIGURA 5.4 - COMPARAÇÃO DOS *WORKLOADS* NOS DOIS PERÍODOS DO DIA COM A MESMA DIVISÃO USUAL DO SISTEMA

5.4 CENÁRIO 2: DIVISÃO GEOGRÁFICA DOS ÁTOMOS ATRAVÉS DA META-HEURÍSTICA BASEADA NO ACO COM MESMO NÚMERO DE PONTOS EM CADA ÁTOMO, NO PERÍODO DAS 10 ÀS 14 HORAS.

No cenário 2, temos a divisão geográfica realizada através da meta-heurística baseada no ACO com os átomos iguais, ou seja, com mesmo número de pontos em cada átomo, conforme a (FIGURA 5.5). Neste cenário os átomos possuem o mesmo número de chamadas, uma vez que as rotas construídas, que chamaremos neste estudo de átomos, são compostas por chamadas emergenciais.

atendidos, é de aproximadamente 0,081; de pelo menos um servidor encontrar-se ocupado é de 0,919.

Os resultados obtidos a partir do modelo para a carga de trabalho dos veículos (*Workload*), o tempo médio de resposta dos veículos e o tempo médio de resposta dos veículos obtidos na amostra encontram-se na (TABELA 5.12).

TABELA 5.12 *WORKLOAD* DOS VEÍCULOS E OS TEMPOS DE RESPOSTA

Veículos	<i>Workload</i>	Modelo	Amostra	Diferença minutos	Diferença %	Desvio Padrão
0	0,45	13,60	12,60	1,00	7,90	0,71
1	0,36	13,60	14,10	-0,50	-3,50	0,35
2	0,33	14,10	14,00	0,10	0,70	0,07
3	0,33	16,40	14,60	1,80	12,30	1,27
4	0,57	14,80	15,00	-0,20	-1,30	0,14
5	0,61	14,20	16,00	-1,80	-11,20	1,27
6	0,65	14,70	19,00	-4,30	-22,60	3,04
7	0,70	15,90	16,00	-0,10	-0,65	0,07
VSG	0,37	14,40	13,83	0,60	4,30	0,42
VSL	0,63	14,90	16,50	-1,60	-8,90	1,13
Total	0,50	14,66	15,16	-0,50	-2,30	0,35

A (TABELA 5.12) apresenta as diferenças em minutos, em porcentagem e também o desvio padrão. Pode-se observar que a diferença mais alta foi encontrada no veículo 6, com 22,6%. Contudo, essa divisão geográfica realizada a partir da meta-heurística apresentou visivelmente uma redução nos tempos de respostas. Essa redução deve ser considerada, pois o objetivo principal do SAE móvel é a redução dos tempos de respostas. Quanto à carga de trabalho dos veículos, constatou-se uma redução tanto nos veículos que atendem serviços graves quanto nos que atendem serviços leves. A comparação está sendo realizada no mesmo intervalo de tempo e com os mesmos veículos em serviço.

A (TABELA 5.13) mostra o tempo de resposta para cada átomo obtido a partir do modelo e da amostra coletada.

TABELA 5.13 - TEMPOS MÉDIO DE RESPOSTA DE CADA ÁTOMO

Átomos	Modelo	Amostra	Diferença minutos	Diferença %	Desvio Padrão
1-1a	12,98	11,00	-1,98	-15,25	1,40
2-1b	12,93	15,41	2,48	19,21	1,76
3-2a	12,73	12,84	0,11	0,89	0,08
4-2b	12,92	14,17	1,25	9,69	0,89
5-3a	14,82	15,30	0,48	3,24	0,34
6-3b	15,29	16,77	1,48	9,65	1,04
7-4a	16,88	14,89	-1,99	-11,82	1,41
8-4b	17,51	15,73	-1,78	-10,17	1,26
Va	12,89	13,36	0,47	3,62	0,33
Vb	16,13	15,67	-0,46	-2,82	0,32
Total	14,51	14,51	0,01	0,04	0,00

Verifica-se, pela (TABELA 5.13), que os tempos fornecidos pelo modelo e pela amostra apresentam diferenças pouco significativas, exceto no átomo 2. Contudo, os resultados obtidos nessa nova divisão geográfica apresentaram menores tempos de respostas quando comparados com os do cenário 1.

Quanto à frequência de despachos para o servidor i se deslocar até o átomo j , o modelo forneceu um valor de 0,54 para a fração dentre todos os despachos que são interáreas de cobertura.

A fração dos despachos que são interáreas (*Backup*) de cobertura de cada um dos veículos pode ser observada na (TABELA 5.14).

TABELA 5.14 - FRAÇÃO DOS DESPACHOS DOS VEÍCULOS QUE SÃO INTERÁREAS

Veículos	<i>Backup</i>
0	0,443
1	0,351
2	0,321
3	0,321
4	0,562
5	0,602
6	0,641
7	0,696
VSG	0,359
VSL	0,632

Os veículos apresentaram índices menores quando comparados aos índices do cenário 1, período da manhã. Também se mantiveram com menor ocupação os veículos destinados a atender casos graves do que os veículos destinados a atender os casos leves. A frequência de despacho de cada veículo em sua área de cobertura primária e a frequência de atendimentos *backup* pode ser observada na (TABELA 5.15).

TABELA 5.15 FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO

Veículos	Cobertura primária	<i>Backup</i>
0	0,435	0,443
1	0,493	0,351
2	0,666	0,321
3	0,498	0,321
4	0,565	0,562
5	0,600	0,602
6	0,540	0,641
7	0,517	0,696
VSG	0,523	0,359
VSL	0,555	0,632

Neste caso também aconteceu um equilíbrio entre os veículos que atendem a chamados graves e os veículos que atendem a chamados leves, quando atendem a chamados dentro da sua cobertura primária.

O modelo forneceu um tempo médio de deslocamentos para chamadas na fila de espera, em minutos, de 14,78. O tempo médio global de resposta do deslocamento, independentemente do átomo gerador do chamado e da equipe que atendeu ao chamado, foi de 14,66 minutos. Na empresa, esse tempo foi calculado em aproximadamente 15,67 minutos, o que caracteriza uma diferença de 1 minuto para menos nesta nova divisão baseada na mudança do algoritmo proposto na tese.

A (TABELA 5.16) apresenta os tempos médios de deslocamento aos átomos do problema abordado no modelo e na amostra. Esses resultados foram obtidos a partir da divisão geográfica baseada no ACO, heurística proposta nesta tese, com átomos com mesmo número de pontos.

O resultado da aplicação da heurística proposta, conforme equação 5.4, mostra que o modelo hipercubo de filas apresentou tempos próximos dos tempos coletados na amostra, validando a metodologia.

TABELA 5.16 TEMPOS MÉDIO DE VIAGEM DE CADA VEÍCULO PARA CADA ÁTOMO

Átomos	Modelo	Amostra	Diferença minutos	Diferença %	Desvio Padrão
1-1a	12,98	13,85	0,87	6,70	0,62
2-1b	12,93	13,85	0,92	7,12	0,65
3-2a	12,73	13,05	0,32	2,51	0,23
4-2b	12,92	13,05	0,13	1,01	0,09
5-3a	14,82	14,10	-0,72	-4,86	0,51
6-3b	15,29	14,10	-1,19	-7,78	0,84
7-4a	16,88	14,25	-2,63	-15,58	1,86
8-4b	17,51	14,25	-3,26	-18,62	2,31
Va	14,35	13,81	-0,54	-3,91	0,38
Vb	14,66	13,81	0,85	6,15	0,60
Total	14,51	13,81	0,70	5,05	0,49

5.5 CENÁRIO 2.1: DIVISÃO GEOGRÁFICA DOS ÁTOMOS ATRAVÉS DA META-HEURÍSTICA BASEADA NO ACO COM ÁTOMOS COM NÚMERO DE PONTOS DIFERENTES, PERÍODO DAS 10 ÀS 14 HORAS.

Neste cenário, a divisão geográfica foi realizada também através da meta-heurística baseada no ACO. A construção desse algoritmo foi parecida com a construção do anterior; contudo, o algoritmo foi modificado de forma a considerar a quantidade diferente de pontos que compõem cada novo átomo. Com o objetivo de diversificar a divisão geográfica e, assim, obter outras medidas de desempenho quando aplicado o modelo hipercubo de filas.

O algoritmo foi modificado diversas vezes, sempre buscando a melhor divisão geográfica. O programa foi executado com 500 iterações, sempre variando os parâmetros taxa de fixação (ρ), o parâmetro que controla a influência do feromônio (α) e o parâmetro que controla a influência da informação heurística (β).

Neste cenário, a probabilidade de se encontrar o sistema vazio, bem como a probabilidade de se encontrar o sistema saturado, com usuários aguardando para serem atendidos e a probabilidade de pelo menos um servidor encontrar-se ocupado foi bem próxima do resultado obtido no cenário 2. Contudo, a nova divisão, (FIGURA 5.6), apresentou medidas de desempenho diversificadas, em alguns casos reduzindo o tempo de resposta.

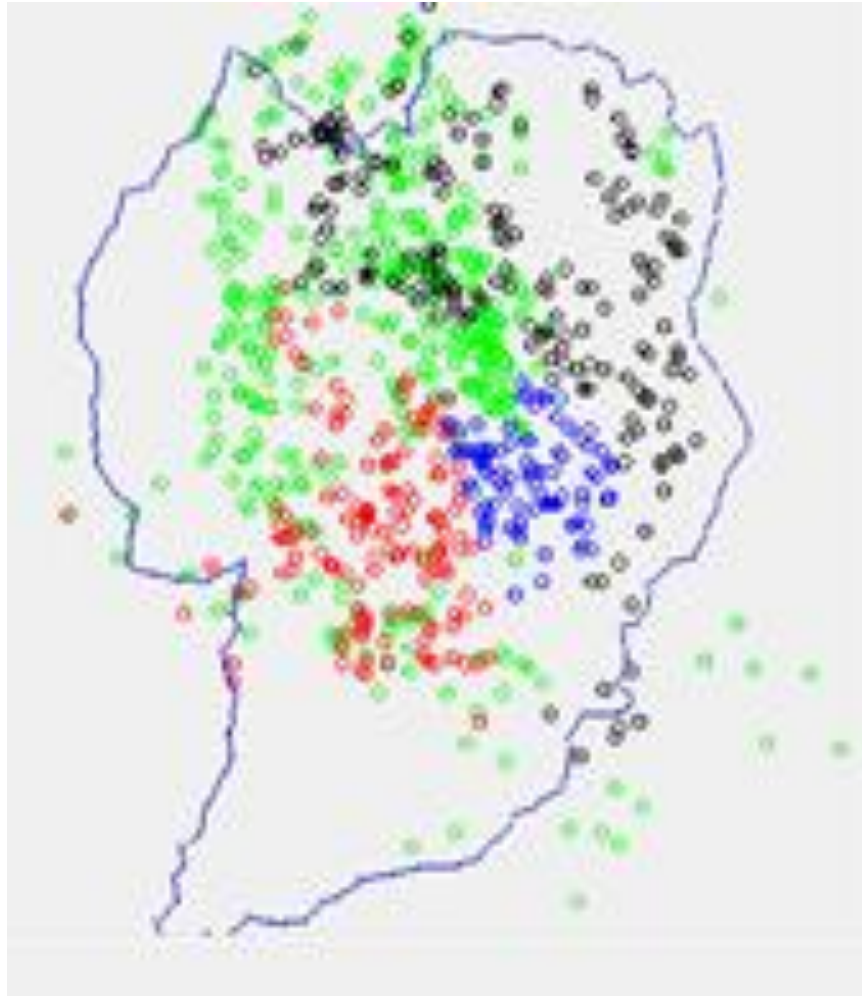


FIGURA 5.6 - DIVISÃO GEOGRÁFICA BASEADA NO ACO COM ÁTOMOS DIFERENTES

Os resultados obtidos a partir do modelo para esse novo cenário 2.1, considerando a carga de trabalho dos veículos (*Workload*), o tempo médio de resposta dos veículos e o tempo médio de resposta dos veículos obtidos na amostra encontram-se na (TABELA 5.17).

TABELA 5.17 - *WORKLOAD* DOS VEÍCULOS E OS TEMPOS DE RESPOSTA

Veículos	<i>Workload</i>	Modelo (minutos)	Amostra (minutos)	Diferença minutos	Diferença %	Desvio Padrão
0	0,422	10,963	12,600	-1,637	-12,994	1,158
1	0,324	12,154	14,100	-1,946	-13,804	1,376
2	0,421	13,370	14,000	-0,630	-4,502	0,446
3	0,336	13,475	14,600	-1,125	-7,707	0,796
4	0,556	13,066	15,000	-1,934	-12,890	1,367
5	0,628	13,646	16,000	-2,354	-14,711	1,664
6	0,674	13,190	19,000	-5,810	-30,577	4,108
7	0,692	13,768	16,000	-2,232	-13,950	1,578
VSG	0,376	12,490	13,825	-1,335	-9,655	0,944
VSL	0,638	13,418	16,500	-3,082	-18,680	2,179
Total	0,507	10,963	12,600	-1,637	-12,994	1,158

A (TABELA 5.17) apresenta para os tempos de resposta, as diferenças em minutos, em porcentagem e também o desvio padrão. Pode-se observar, neste novo cenário, uma redução nos tempos de respostas, tanto para os veículos que atendem casos graves como para os veículos que atendem casos leves.

A (TABELA 5.18) mostra o tempo de resposta para cada átomo obtido a partir do modelo e da amostra coletada.

TABELA 5.18 - TEMPOS MÉDIO DE RESPOSTA DE CADA ÁTOMO

Átomos	Modelo (minutos)	Amostra (minutos)	Diferença minutos	Diferença %	Desvio Padrão
1-1A	12,911	13,310	-0,399	-2,995	0,282
2-1B	12,148	16,710	-4,562	-27,303	3,226
3-2A	9,440	14,070	-4,630	-32,909	3,274
4-2B	11,478	15,540	-4,062	-26,141	2,873
5-3A	13,523	13,100	0,423	3,226	0,299
6-3B	13,905	16,870	-2,965	-17,578	2,097
7-4A	13,591	16,350	-2,759	-16,876	1,951
8-4B	14,639	16,790	-2,151	-12,813	1,521
VA	12,366	14,208	-1,841	-12,961	1,302
VB	13,042	16,470	-3,428	-20,813	2,424
Total	12,704	15,339	-2,635	-17,178	1,863

Verifica-se, pela (TABELA 5.18), que os tempos fornecidos pelo modelo e pela amostra apresentam diferenças significativas. Essa nova divisão apresentou uma redução média de 12,96% de tempo para o atendimento de casos graves e uma redução de 20,8% para casos considerados leves, o que resulta em uma média global de 17,18% de redução no tempo de atendimento.

Quanto à frequência de despachos para o servidor i ir até o átomo j , o modelo forneceu um valor de 0,540 para a fração dentre todos os despachos que são interáreas de cobertura.

A fração dos despachos que são interáreas (*Backup*) de cobertura de cada um dos veículos pode ser observada na (TABELA 5.19)

TABELA 5.19 - FRAÇÃO DOS DESPACHOS DOS VEÍCULOS QUE SÃO INTERÁREAS

Veículos	<i>Backup</i>
0	0,410
1	0,313
2	0,409
3	0,324
4	0,545
5	0,617
6	0,663
7	0,681
VSG	0,364
VSL	0,628

Mantiveram-se com menor ocupação os veículos destinados a atender casos graves em relação aos veículos destinados a atender os casos leves. A frequência de despacho de cada veículo em sua área de cobertura primária e a frequência de atendimentos *backup* pode ser observada na (TABELA 5.20).

TABELA 5.20 - FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO

Veículos	Cobertura primária	<i>Backup</i>
0	0,465	0,410
1	0,553	0,313
2	0,455	0,409
3	0,597	0,324
4	0,556	0,545
5	0,551	0,617
6	0,525	0,663
7	0,590	0,681
VSG	0,517	0,364
VSL	0,555	0,628

Neste caso também aconteceu um equilíbrio entre a quantidade de veículos que atendem a chamados graves e os veículos que atendem a chamados leves, quando atenderem a área de cobertura primária.

O modelo forneceu um tempo médio de deslocamentos para chamadas na fila de espera, em minutos, de 12,33. O tempo médio global de resposta do deslocamento, independentemente do átomo gerador do chamado e da equipe que atendeu ao chamado foi de 12,77 minutos. Na empresa, esse tempo foi calculado em aproximadamente 15,67 minutos, o que caracteriza uma diferença de 2,89 minutos para menos, ou seja, uma redução significativa de 18% considerando a nova divisão geográfica proposta neste trabalho.

A (TABELA 5.21) apresenta os tempos médios de viagem para cada átomo do problema abordado, no modelo e na amostra.

TABELA 5.21 - TEMPOS MÉDIO DE VIAGEM DE CADA VEÍCULO PARA CADA ÁTOMO

Átomos	Modelo	Amostra	Diferença minutos	Diferença %	Desvio Padrão
1-1a	12,91	15,38	-2,46	-16,02	1,74
2-1b	12,15	15,38	-3,23	-20,99	2,28
3-2a	9,44	13,93	-4,49	-32,21	3,17
4-2b	11,48	13,93	-2,45	-17,58	1,73
5-3a	13,52	15,31	-1,79	-11,69	1,27
6-3b	13,90	15,31	-1,41	-9,20	1,00
7-4a	13,59	16,20	-2,61	-16,11	1,85
8-4b	14,64	16,20	-1,56	-9,64	1,10
Va	12,37	15,20	-2,84	-18,66	2,01
Vb	13,04	15,20	-2,16	-14,22	1,53
Total	12,70	15,20	-2,50	-16,44	1,77

Neste novo cenário, com a nova distribuição espacial dos átomos, a diferença nos tempos de respostas mostrou-se satisfatória. A redução média nos tempos de resposta foi de 16,44%.

5.6 COMPARAÇÃO DAS MEDIDAS DE DESEMPENHO DOS CENÁRIOS 1, 2 E 2.1

Comparando os três cenários propostos neste estudo – a divisão usual (BPM), a divisão baseada no ACO com átomos com mesma quantidade de pontos e

a divisão também baseada na meta-heurística ACO com átomos com quantidade de pontos diferentes, no período da manhã –, conclui-se, conforme os dados apresentados na (TABELA 5.22), um melhor resultado no cenário 2. A tabela mostra a comparação das *workloads*.

TABELA 5.22 - COMPARAÇÃO DAS *WORKLOADS* NAS TRÊS DIVISÕES GEOGRÁFICAS

Veículos	<i>Workload</i> BPM	<i>Workload</i> ACO - Átomos iguais	<i>Workload</i> ACO- Átomos diferentes
0	0,42	0,45	0,42
1	0,51	0,36	0,32
2	0,49	0,33	0,42
3	0,31	0,33	0,34
4	0,60	0,57	0,56
5	0,65	0,61	0,63
6	0,74	0,65	0,67
7	0,68	0,70	0,69
VG	0,43	0,37	0,38
VL	0,67	0,63	0,64
Total	0,55	0,50	0,51

A carga de trabalho dos veículos, uma importante medida de desempenho do ponto de vista do gerente do sistema, mostrou-se mais adequada ao SAE móvel com as divisões geográficas efetuadas através da meta-heurística baseada no ACO. Os resultados obtidos para a medida de desempenho da carga de trabalho dos veículos, *workload*, apresentaram uma redução e uma melhor distribuição das cargas entre os veículos, principalmente para os veículos que atendem preferencialmente os casos graves. Nos dois cenários propostos, houve uma redução de 16% das *workloads* para os veículos que atendem casos graves e de 6,3% para os veículos que atendem casos leves. Essa redução é de grande importância para o sistema, uma vez que o ideal é que esses veículos fiquem desocupados o maior tempo possível, devido às características de aleatoriedade do SAE móvel.

A (TABELA 5.23) apresenta a comparação dos tempos de resposta fornecidos pelo modelo hipercubo no horário de pico, das 10 às 14 horas, com a divisão geográfica usual e com a divisão baseada no ACO com átomos iguais.

TABELA 5.23 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS

Veículos	Tempo de resposta-BPM (minutos)	Tempo de resposta-ACO Átomos iguais (minutos)	Diferença minutos	Diferença %
0	14,20	13,60	0,60	4,41
1	16,10	13,60	2,50	18,38
2	15,40	14,10	1,30	9,22
3	15,60	16,40	-0,80	-4,88
4	15,50	14,80	0,70	4,73
5	15,10	14,20	0,90	6,34
6	16,40	14,70	1,70	11,57
7	16,40	15,90	0,50	3,15
VG	15,33	14,40	0,93	6,46
VL	15,85	14,90	0,95	6,38
Total	15,59	14,66	0,93	6,34

O tempo de resposta é uma medida de desempenho importante do ponto de vista do usuário do sistema. Conforme se pode observar na TABELA 5.23, o menor valor foi obtido com a utilização da metodologia proposta, baseada no ACO com átomos de mesmo tamanho. A redução foi em média superior a 6,3% para os dois veículos de atendimento emergencial, VG e VL.

A TABELA 5.24 apresenta a comparação dos tempos de resposta para as duas divisões geográficas. A primeira divisão, a usual (BPM), e a segunda divisão geográfica através da meta-heurística baseada no ACO, com átomos de tamanhos diferentes.

TABELA 5.24 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS

Veículos	Tempo de resposta - BPM (minutos)	Tempo de resposta ACO Átomos diferentes (minutos)	Diferença Tempo minutos	Diferença Tempo %
0	14,20	10,96	3,24	29,53
1	16,10	12,15	3,95	32,47
2	15,40	13,37	2,03	15,19
3	15,60	13,47	2,13	15,77
4	15,50	13,07	2,43	18,62
5	15,10	13,65	1,45	10,65
6	16,40	13,19	3,21	24,33
7	16,40	13,77	2,63	19,12
VG	15,33	12,49	2,84	22,74
VL	15,85	13,42	2,43	18,13
Total	15,59	12,95	2,64	20,35

Novamente, pode-se observar na (TABELA 5.24) uma visível redução nos tempos de respostas com a divisão geográfica, baseada no ACO, proposta neste trabalho. A redução em média foi de 20,35%, quase 3 minutos de redução nos tempos para os veículos que atendem prioritariamente casos graves.

Finalmente, na (TABELA 5.25) encontra-se a comparação dos tempos de resposta com as duas divisões geográficas baseadas no ACO. Na primeira, os átomos possuem o mesmo número de pontos; na segunda, os átomos possuem número de pontos diferenciados.

TABELA 5.25 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS

Veículos	Tempo de resposta ACO - Átomos iguais (minutos)	Tempo de resposta ACO - Átomos diferentes (minutos)	Diferença minutos	Diferença %
0	13,60	10,96	2,64	24,06
1	13,60	12,15	1,45	11,90
2	14,10	13,37	0,73	5,46
3	16,40	13,47	2,93	21,71
4	14,80	13,07	1,73	13,27
5	14,20	13,65	0,55	4,06
6	14,70	13,19	1,51	11,45
7	15,90	13,77	2,13	15,49
VG	14,40	12,49	1,91	15,29
VL	14,90	13,42	1,48	11,05
Total	14,66	12,95	1,71	13,17

A comparação realizada entre as divisões geográficas propostas neste trabalho – mudança no algoritmo baseado na meta-heurística ACO – mostrou nitidamente a eficiência da proposta para a divisão com átomos de tamanhos diferentes. Houve uma redução nos tempos de mais de 15,29% nos veículos que atendem casos graves e de 11% para veículos que atendem preferencialmente casos leves. Essa redução é de extrema importância para os SAE móveis.

Outra redução significativa foi a do tempo de resposta dos veículos aos novos átomos propostos. Podemos verificar, pela (TABELA 5.26), a comparação nos dois cenários distintos: divisão usual e divisão baseada no ACO com átomos de mesmo tamanho.

TABELA 5.26 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS NAS DUAS DIVISÕES GEOGRÁFICAS

Átomos	Modelo (BPM)	Modelo ACO - átomos iguais	Diferença minutos	Diferença %	Desvio Padrão
1-12A-1A	14,60	12,98	1,62	12,48	1,15
2-12B-1B	14,50	12,93	1,57	12,14	1,11
3-13A-2A	13,90	12,73	1,17	9,19	0,83
4-13B-2B	14,20	12,92	1,28	9,91	0,91
5-20A-3A	16,60	14,82	1,78	12,01	1,26
6-20B-3B	17,40	15,29	2,11	13,80	1,49
7-23A-4A	16,70	16,88	-0,18	-1,07	0,13
8-23B-4B	17,10	17,51	-0,41	-2,34	0,29
VA	15,45	14,35	1,10	7,67	0,78
VB	15,80	14,66	1,14	7,78	0,81
Total	15,60	14,51	1,09	7,53	0,77

De acordo com os dados apresentados na (TABELA 5.26) obteve-se uma redução dos tempos de resposta aos novos átomos de 7,53% quando comparados aos tempos da divisão segundo o BPM, tanto nos casos graves quanto nos casos leves.

A (TABELA 5.27) apresenta a comparação dos tempos de resposta utilizando divisão usual e a divisão baseada no ACO com átomos de tamanhos diferentes.

TABELA 5.27 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS NAS DUAS DIVISÕES GEOGRÁFICAS

Átomos	Modelo (BPM)	Modelo ACO átomos diferentes	Diferença minutos	Diferença %	Desvio Padrão
1-12A-1A	14,60	12,91	1,69	13,08	1,19
2-12B-1B	14,50	12,15	2,35	19,36	1,66
3-13A-2A	13,90	9,44	4,46	47,25	3,15
4-13B-2B	14,20	11,48	2,72	23,72	1,93
5-20A-3A	16,60	13,52	3,08	22,76	2,18
6-20B-3B	17,40	13,90	3,50	25,14	2,47
7-23A-4A	16,70	13,59	3,11	22,88	2,20
8-23B-4B	17,10	14,64	2,46	16,81	1,74
VA	15,45	12,37	3,08	24,94	2,18
VB	15,80	13,04	2,76	21,15	1,95
Total	15,60	12,70	2,90	22,80	2,05

A comparação dos tempos de resposta em cada átomo geográfico, conforme apresentado na (TABELA 5.27), mostrou que a divisão geográfica baseada no ACO

com átomos de tamanhos diferentes reduz em 22,80% o tempo de resposta em média.

A (TABELA 5.28) apresenta a comparação dos tempos de resposta nos dois cenários propostos baseados no ACO.

TABELA 5.28 - COMPARAÇÃO DOS TEMPOS DE RESPOSTAS NAS DUAS DIVISÕES GEOGRÁFICAS

Átomos	Modelo-ACO átomos iguais	Modelo-ACO átomos diferentes	Diferença (min)	Diferença %	Desvio Padrão
1-12A-1A	12,98	12,91	0,07	0,53	0,05
2-12B-1B	12,93	12,15	0,78	6,44	0,55
3-13A-2A	12,73	9,44	3,29	34,86	2,33
4-13B-2B	12,92	11,48	1,44	12,57	1,02
5-20A-3A	14,82	13,52	1,30	9,60	0,92
6-20B-3B	15,29	13,90	1,39	9,96	0,98
7-23A-4A	16,88	13,59	3,29	24,20	2,33
8-23B-4B	17,51	14,64	2,87	19,62	2,03
VA	14,35	12,37	1,98	16,04	1,40
VB	14,66	13,04	1,62	12,41	1,14
Total	14,51	12,70	1,80	14,20	1,28

De acordo com os dados expostos na (TABELA 5.28), conclui-se que o algoritmo implementado neste estudo baseado no ACO, a fim de realizar divisões geográficas com átomos de tamanhos diferentes, também apresentou resultados superiores com relação aos tempos de resposta quando comparados aos dos outros cenários. Neste caso, a redução de tempo foi ainda maior para atendimentos graves, em média 16% de redução, quase 2 minutos. Fator de extrema relevância para o SAE móvel.

Quanto à frequência de despachos para o servidor i ir até o átomo j , o valor reduziu de 0,582 para 0,541 a partir da nova divisão baseada na meta-heurística, para a fração dentre todos os despachos que são interáreas de cobertura.

A fração dos despachos que são interáreas (*Backup*) de cobertura de cada um dos veículos pode ser visualizada na (TABELA 5.29).

Constata-se uma redução na taxa de ocupação dos veículos, principalmente na dos veículos que atendem os casos graves, de 16,16%, da taxa de atendimento quando solicitados como *backup* conforme mostra a (TABELA 5.29).

TABELA 5.29 COMPARAÇÃO DOS BACKUPS NOS DOIS CENÁRIOS

Veículos	Backup (BPM)	Backup (ACO)	Desvio %
0	0,40	0,44	-9,48
1	0,49	0,35	40,46
2	0,48	0,32	48,91
3	0,30	0,32	-7,48
4	0,57	0,56	1,96
5	0,62	0,60	3,32
6	0,72	0,64	12,32
7	0,66	0,70	-5,32
VSG	0,42	0,36	16,16
VSL	0,64	0,63	1,74

A comparação – utilizando divisão usual e a proposta baseada na meta-heurística – da frequência de despacho de cada veículo, ou seja, da frequência de atendimentos em sua área de cobertura primária e da frequência de atendimentos *backup* pode ser vista na (TABELA 5.30).

TABELA 5.30 - COMPARAÇÃO NA FREQUÊNCIA DE DESPACHO PARA CADA VEÍCULO

Veículos	Cobertura primária (BPM)	Backup (BPM)	Cobertura primária (ACO)	Backup (ACO)
0	0,533	0,401	0,435	0,443
1	0,478	0,493	0,493	0,351
2	0,574	0,478	0,666	0,321
3	0,785	0,297	0,498	0,321
4	0,520	0,573	0,565	0,562
5	0,577	0,622	0,600	0,602
6	0,390	0,720	0,540	0,641
7	0,548	0,359	0,517	0,696
VSG	0,592	0,417	0,523	0,359
VSL	0,508	0,568	0,555	0,632

Conforme se pode visualizar na (TABELA 5.30), houve uma redução nas cargas de trabalho dos veículos que atendem prioritariamente casos graves quando aplicada a nova metodologia proposta.

Com relação ao tempo médio de deslocamentos para chamadas na fila de espera, a nova metodologia mostrou que houve uma redução considerável tanto no tempo de espera quanto no tempo médio global de resposta do deslocamento,

independentemente do átomo gerador do chamado e da equipe que atendeu ao chamado, conforme a (TABELA 5.31).

TABELA 5.31 - COMPARAÇÃO DOS TEMPOS DE ESPERA NA FILA E TEMPO MÉDIO GLOBAL NOS TRÊS CENÁRIOS

Tempo (minutos)	Cenário 1	Cenário 2	Cenário 2.1	Diferença% Cenários 1 e 2	Diferença% Cenários 1 e 2.1
Espera na fila	15,61	14,68	12,33	6,34	26,60
Médio Global	15,65	14,66	12,77	6,75	22,55
Amostra	15,67	15,67	15,67	-	-

Os tempos de espera na fila e o tempo médio global tiveram uma redução média de 6% na comparação do cenário usual em relação à divisão geográfica baseada na meta-heurística ACO com átomos iguais. Contudo, a redução significativa ocorreu na comparação do cenário usual em relação à divisão geográfica baseada na meta-heurística ACO com átomos diferentes: na fila de espera, a redução dos tempos foi de 26,60% e no tempo médio global, próxima de 22,6%.

Essa redução nos tempos de espera comprova nitidamente que a proposta de modificação no algoritmo baseado no ACO aplicada junto ao modelo hipercubo de filas, fornece medidas de desempenho importantes para o sistema, contribuindo para uma tomada de decisão mais acertada por parte do gerente do SAE móvel; afinal, a redução nos tempos de respostas é um dos fatores mais relevantes no SAE móvel.

O tempo médio de viagem de cada veículo ao átomo também foi analisado sob o ponto de vista dos cenários propostos.

A (TABELA 5.32) apresenta a comparação dos tempos médios de viagem para cada átomo bem como suas diferenças e o desvio padrão.

TABELA 5.32 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO

Átomos	Modelo (BPM)	Modelo (ACO) átomos iguais	Diferença (minutos)	Diferença %	Desvio Padrão
1-12A	13,93	12,98	0,95	7,32	0,67
2-12B	16,50	12,93	3,57	27,61	2,52
3-13A	14,63	12,73	1,90	14,93	1,34
4-13B	16,72	12,92	3,80	29,41	2,69
5-20A	14,50	14,82	-0,32	-2,16	0,23
6-20B	14,10	15,29	-1,19	-7,78	0,84
7-23A	17,40	16,88	0,52	3,08	0,37
8-23B	17,06	17,51	-0,45	-2,57	0,32
VA	15,11	14,35	0,76	5,30	0,54
VB	16,09	14,66	1,43	9,75	1,01
Total	15,60	14,51	1,09	7,53	0,77

De maneira geral, os tempos médios de viagem sofreram uma significativa redução, em média 7,53 %, quando se trabalhou com a aplicação do ACO com átomos de mesmo tamanho.

A (TABELA 5.33) apresenta a comparação dos tempos médios de viagem para cada átomo na divisão usual e na divisão geográfica baseada no ACO com átomos de tamanhos diferentes.

TABELA 5.33 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO

Átomos	Modelo Cenário 1	Modelo Cenário 2.1	Diferença (minutos)	Diferença %	Desvio Padrão
1-12A	13,93	12,91	1,02	7,89	0,72
2-12B	16,50	12,15	4,35	35,83	3,08
3-13A	14,63	9,44	5,19	54,98	3,67
4-13B	16,72	11,48	5,24	45,67	3,71
5-20A	14,50	13,52	0,98	7,23	0,69
6-20B	14,10	13,90	0,20	1,41	0,14
7-23A	17,40	13,59	3,81	28,03	2,69
8-23B	17,06	14,64	2,42	16,54	1,71
VA	15,11	12,37	2,74	22,19	1,94
VB	16,09	13,04	3,05	23,37	2,16
Total	15,60	12,70	2,90	22,80	2,05

Neste contexto, comparando a divisão geográfica usual, cenário 1, com a divisão heurística baseada no ACO, cenário 2.1, constata-se que houve uma redução nos tempos de viagem em média de 22,80%.

Na comparação dos tempos médios de viagem das duas divisões heurísticas baseadas no ACO, conforme apresentados na TABELA 5.34, verifica-se que a divisão realizada com átomos de tamanhos diferentes apresentou melhores resultados: redução de aproximadamente 1,8 minutos.

TABELA 5.34 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO

Átomos	Modelo Cenário 2	Modelo cenário 2.1	Diferença minutos	Diferença %	Desvio Padrão
1-12A	12,98	12,91	0,07	0,53	0,05
2-12B	12,93	12,15	0,78	6,44	0,55
3-13A	12,73	9,44	3,29	34,86	2,33
4-13B	12,92	11,48	1,44	12,57	1,02
5-20A	14,82	13,52	1,30	9,60	0,92
6-20B	15,29	13,90	1,39	9,96	0,98
7-23A	16,88	13,59	3,29	24,20	2,33
8-23B	17,51	14,64	2,87	19,62	2,03
VA	14,35	12,37	1,98	16,04	1,40
VB	14,66	13,04	1,62	12,41	1,14
Total	14,51	12,70	1,80	14,20	1,28

Conforme a (TABELA 5.35), nos três cenários propostos para o período da manhã, a divisão heurística, baseada no ACO com átomos com número de pontos distintos apresentou resultados com maior redução nos tempos médios de viagem.

TABELA 5.35 - COMPARAÇÃO DOS TEMPOS MÉDIOS DE VIAGEM PARA CADA ÁTOMO NOS TRÊS CENÁRIOS PROPOSTOS

Átomos	Modelo (BPM)	Modelo (ACO) átomos iguais	Modelo (ACO) átomos diferentes
1-12A	13,93	12,98	12,91
2-12B	16,50	12,93	12,15
3-13A	14,63	12,73	9,44
4-13B	16,72	12,92	11,48
5-20A	14,50	14,82	13,52
6-20B	14,10	15,29	13,90
7-23A	17,40	16,88	13,59
8-23B	17,06	17,51	14,64
VA	15,11	14,35	12,37
VB	16,09	14,66	13,04
Total	15,60	14,51	12,70

Logo, conclui-se que a melhor divisão para o SAE móvel, com menores tempos de respostas, melhores distribuições das *workload* e frequências mais bem distribuídas é a divisão realizada conforme o cenário 2.1.

6 CONCLUSÕES E PROPOSTAS DE MELHORAMENTO

Este trabalho teve como objetivo analisar o Sistema Atendimento Emergencial (SAE) móvel, através do modelo matemático hipercubo de filas e de um novo algoritmo construído, baseado no *Ant Colony Optimization* (ACO).

O modelo hipercubo de filas é descritivo, e se aplicado isoladamente, não permite a solução direta de problemas de localização. Não se enquadra como um modelo de otimização. Contudo, é capaz de fornecer importantes medidas de desempenho para avaliação do SAE móvel. Foi utilizado inicialmente no patrulhamento policial em Nova Iorque, com a finalidade de avaliar diversas medidas de desempenho. As medidas são fornecidas de acordo com cada cenário sugerido pelo avaliador do sistema. O hipercubo de filas mostrou-se uma ferramenta muito eficiente e eficaz.

A outra metodologia utilizada baseada no algoritmo ACO, consiste basicamente num algoritmo guloso de probabilidade. Foi desenvolvido inicialmente por Dorigo *et al* (1996) e tinha como principal objetivo resolver o PCV. Contudo, a eficiência do ACO foi conhecida com sua aplicação em problemas *job-shop scheduling* e roteamento de veículos.

Com o objetivo de criar uma metodologia de avaliação para o SAE móvel, o modelo hipercubo de filas com prioridade foi aplicado em um estudo de caso, em uma empresa particular na cidade de Curitiba, PR, com a finalidade inicial de validar as 9 hipóteses do modelo, condições necessárias para utilização da ferramenta matemática.

Uma das hipóteses do modelo diz respeito a divisão da região em átomos geográficos independentes. A região em estudo deverá ser passível de ser dividida. Inicialmente a divisão da região foi a usual no sistema, com quatro átomos independentes segundo o critério do Batalhão da Polícia Militar (BPM) de Curitiba. Dentro de cada átomo houve uma subdivisão de acordo com as prioridades da empresa, atendimentos de casos graves e não tão graves. Cabe ressaltar que essa prioridade foi atendida perfeitamente pelo modelo hipercubo de filas.

Posteriormente, a divisão geográfica foi realizada através da construção de um novo algoritmo baseado no ACO. Essa segunda divisão aconteceu de duas formas distintas, a primeira com átomos com igual número de pontos e, posteriormente, com átomos com número de pontos diferentes.

A coleta de dados aconteceu em dois períodos distintos ao longo do dia, um no período da manhã (das 10h às 14h) e outro no período da noite (18h às 23h). A análise desses períodos distintos reforçou a importância de se considerar diferentes configurações do SAE móvel ao longo do dia. Observou-se que, apesar do período da tarde ser o período com maior número de chamados, foi o período da manhã que se mostrou mais desfavorável, apresentando o maior congestionamento, conforme pode-se observar através de algumas medidas de desempenho.

Após a coleta e análise de dados, o modelo hipercubo foi aplicado a diversos cenários. O primeiro cenário diz respeito aos dois períodos de pico e com a divisão usual, depois o modelo foi aplicado apenas no período da manhã com a divisão baseada no Algoritmo ACO. Inicialmente os átomos tinham o mesmo número de pontos e depois os átomos foram construídos com número de pontos distintos. Para cada um desses cenários foram calculadas diversas medidas de desempenho tais como: *Workload*, frequência de despacho e tempo de resposta. Algumas medidas obtidas pelo modelo puderam ser comparadas com os resultados obtidos na amostra, apresentando, na maioria dos resultados, desvios pouco significativos.

Para a aplicação do modelo, devem-se considerar as características específicas do serviço prestado, bem como demanda local, configuração geográfica da cidade e região metropolitana, características dos chamados, sempre levando em consideração os diferentes tipos de atendimento. Nesse estudo de caso, tratou-se apenas de casos de urgência e emergência.

6.1 CONCLUSÕES

Diante de uma população que aumentou consideravelmente nos últimos anos e com perspectivas de crescimento ainda maiores para as próximas décadas, faz-se necessária uma análise no serviço de atendimento móvel a fim de garantir as devidas melhorias e consequentemente a satisfação para todos os usuários desse sistema.

Os altos níveis da demanda de usuários são geralmente incompatíveis com o baixo número de veículos disponíveis em todo SAE móvel do Brasil (TAKEDA, 2000), justificando dessa forma uma análise com propostas de melhorias para todo o sistema.

O problema estudado apresentou um cenário com alto índice de demanda e poucos veículos à disposição para supri-la em determinados períodos críticos do dia (períodos de pico). Esse período crítico pode acontecer mais de uma vez ao longo do dia. No caso do sistema analisado, constatou-se que esse “pico” ocorre tanto no período da manhã como no período da tarde. Com isso, ambos os períodos foram analisados através do modelo proposto. Constatou-se também a necessidade de um melhor balanceamento nas cargas de trabalho dos veículos. Contudo, não basta apenas aumentar o número de veículos para elevar a qualidade do serviço prestado. Necessita-se quantificar essas melhorias por meio de índices de desempenho, o que não é um processo fácil. Com isso justificam-se as investigações e análises científicas do atual sistema de operações e possíveis ajustes através das medidas de desempenho fornecidas pelo modelo, com o objetivo de melhorar o atendimento aos usuários do sistema.

A fim de aplicar o modelo, inicialmente verificou-se todas as hipóteses do modelo hipercubo de filas, conforme apresentado na seção 3, cujo resultado mostrou que o sistema pode ser modelado por essa ferramenta, apresentando desvios pouco significativos quando comparados com os dados reais.

As medidas de desempenho do SAE móvel foram obtidas através do modelo hipercubo de duas formas distintas e em dois períodos do dia, conforme já mencionado. Os resultados obtidos através da meta-heurística baseada no ACO mostram melhores resultados de maneira geral. Houve uma significativa redução nos tempos. A adaptação proposta no algoritmo ACO mostrou-se muito eficiente nos resultados das medidas de desempenho. Destaca-se, por exemplo, para o período da manhã, em que o tempo médio global de deslocamento dos veículos no modelo apresentou um valor igual a 15,65 minutos através da divisão pelo BPM, enquanto que na divisão pelo algoritmo baseado em ACO, átomos iguais, este resultado caiu para 14,66 minutos, o que caracteriza uma redução no tempo de quase 7%. Essa redução é maior ainda quando comparada com o algoritmo baseado em ACO, com átomos com número de pontos diferentes, pois o tempo é de 12,77 minutos, ou seja, 22,5% menor quando comparado com a divisão usual. Outra redução significativa foi observada nas *workloads*, que em média apresentaram 6% de redução, mostrando um melhor balanceamento entre as cargas de trabalho dos veículos.

As melhorias propostas na divisão dos átomos geográficos, obtidas a partir do algoritmo baseado na meta-heurística ACO, foram significativas. Pode-se comprovar que a heurística apresenta-se bem adequada ao problema.

6.2 PROPOSTAS DE MELHORAMENTO

Alguns questionamentos e perspectivas surgiram durante o desenvolvimento deste trabalho. O SAE móvel necessita de melhorias, conforme mencionou-se no decorrer dessa tese e aqui deixamo-nas como sugestões para futuras pesquisas.

Sugerem-se outros estudos híbridos envolvendo técnicas de otimização juntamente com o Modelo Hipercubo de Filas. Por exemplo, a criação de uma nova meta-heurística, baseada na nuvem de partículas, meta-heurística inspirada no voo dos pássaros, a fim de criar outras divisões geográficas e aplicá-la juntamente com o Modelo Hipercubo de Filas adequando a prioridade.

Outra possibilidade de pesquisa seria a de inserir os atendimentos de remoção e consulta na análise, uma vez que representam uma parcela relevante do total de atendimentos do SAE móvel. Entraria como uma terceira classe de usuários. Contudo, deve-se levar em consideração a quantidade de veículos que atendem a esses chamados, uma vez que o número de equações do modelo está interligado com o número de veículos.

É possível acrescentar ao modelo uma restrição quanto ao número de veículos em diferentes períodos do dia, mantendo dessa forma um determinado número de medidas de desempenho em um valor pré-estipulado. Poderia-se, por exemplo, determinar um número de veículos para o período de pico da manhã, com tempo de atendimento inferior a 10 minutos.

Qualquer que seja a melhoria efetiva no SAE móvel, muito acrescentará ao sistema como um todo. A pesquisa operacional é uma área promissora, e acredita-se que há muito ainda a se explorar.

REFERÊNCIAS

- ALBINO, J. C. C. **Quantificação e Localização de Unidades Móveis de Atendimento de Emergência a Interrupções em Redes de Distribuição de Energia Elétrica: Aplicação do Modelo Hipercubo**. Dissertação (Mestrado em engenharia da produção) – Universidade Federal de Santa Catarina, Florianópolis, 1994.
- AUGERAT, P.; BELENGUER, J.; BENAVENT, E.; CORBERÁN, A.; NADDEF, D. ***Separating Capacity Constraints in the CVRP Using Tabu Search***. European Journal of Operations Research, 106 (2-3), p 546-557, 1998.
- BABA, C. M.; CORRÊA, F. H. M.; SOUZA, S. A. C.; WAHBA, T. M. **Programação e Roteirização de Veículos para o Transporte de Pessoas Portadoras de Deficiência**. Centro Universitário do Instituto Mauá de Tecnologia, São Caetano do Sul, 2004a.
- BABA, C. M.; CORRÊA, F. H. M.; SOUZA, S. A. C.; WAHBA, T. M.; MEDINA, A. C. **Otimização da Colônia de Formigas Aplicada ao Problema da Programação e Roteirização de Veículos para o Transporte de Pessoas Portadoras de Deficiência**. XXIV ENEGEP, Florianópolis, SC, p. 2966 - 2973, 2004b.
- BATTA, R.; DOLAN, J. M.; KRISHNAMURTHY, N. M. ***The Maximal Expected Covering Location Problem***. Revised, Transportation Science 23, p. 277- 287, 1989.
- BELTRAMI, E. J. ***Models for Public Systems Analysis***. New York, Academic Press, 1977.
- BONABEAU, E.; DORIGO, M.; THÉRAULAZ, G. ***Swarm Intelligence: From Natural to Artificial Systems***. Oxford, UK. Oxford University Press, 1999.
- BURWELL, T. H.; JARVIS, J. P.; MCKNEW, M. A. ***Modeling Co-located Servers and Dispatch Ties the Hypercube Model***. Journal of computers and operations research, v.20, n.2, p. 113-119, 1993.
- CAMELO, G. R.; COELHO, A. S.; BORGES, R. M.; SOUZA, R. M. **Teoria de Filas e da Simulação Aplicada ao Embarque de Minério de ferro e Manganês no Terminal Marítimo de Ponta da Madeira**. Universidade do Estado do Rio de Janeiro, RJ, vol 29, p 1-16, 2010.
- CARVALHO, M. B.; YAMAKAMI, A. **Meta-heurística Híbrida de Sistema de Colônia de Formigas e Algoritmo Genético para o problema do Caixeiro Viajante**. Tendências em matemática Aplicada e Computacional. V.9, n.1, 2008.
- CHAIKEN, J. M. ***Hypercube Queuing Model: Executive Summary***. New York: The Rand Corporation, 1975, p. 17. Disponível em: <http://mtc-m17.sid.inpe.br/col/sid.inpe.br/mtc-17@80/2007/12.07.11.42/doc/publicacao.pdf>. Acesso abril 2013.

CHAIKEN, J. M. ***Transfer of Emergency Service Deployment Models to Operating Agencies***. Management Science, v 24, p. 719-731, 1978.

CHIYOSHI F.; GALVÃO R. D. ***A statistical analisys of simulated annealing applied to the p-median problem***. Annals of Operational Research 96, p. 61-74, 2000.

CHIYOSHI F.; GALVÃO R.D., MORABITO R. ***O uso do modelo hipercubo na solução de problemas de localização probabilísticos***. Gestão & Produção 7(2), 146- 174, 2000.

CHIYOSHI F., GALVÃO R. D., MORABITO R. ***Modelo Hipercubo: Análise e Resultados para o Caso de Servidores não Homogêneos***. Pesquisa Operacional, vol 21, n.2, p. 199-218, 2001.

CHIYOSHI F., GALVÃO R. D., MORABITO R. ***A note on solution to the maximal expected covering location problem***. Computers and Operations Research 30 (1), p. 87-96, 2003.

COSTA, D. M. B. ***Uma Metodologia Iterativa para Determinação de Zonas de Atendimento de Serviços Emergenciais***. Tese (Doutorado em engenharia de produção) - Departamento de Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis, Florianópolis, 2003.

DENEUBOURG, J. L.; ARON, S.; GOSS, S.; PASTEELS, J. M. ***The self-organizing exploratory pattern of the argentine ant***. Journal of Insect Behavior, 3, 159-168, 1990.

DONATI, A. V.; MONTEMANNI, R.; CASAGRANDE, N.; RIZZOLI, A. E.; GAMBARDELLA, L. M. ***Time Dependent Vehicle Routing Problem with a Multi Ant Colony System***. Instituto Dalle Molle di Studi sull'intelligenza Artificiale (IDSIA) Galleria 2, 6928 Manno, Switzerland, 2008.

DORIGO, M. ***Otimization, learning and natural algorithms***. Tese de doutorado. Politecnico di Milano, Itália, 1992.

DORIGO, M.; GAMBARDELLA, L. M. ***Ant Colony System: a Cooperative Learning Approach to the Traveling Salesman Problem***. IEEE Trnsations on Evolutionaty Computation, 1 (1):53-66, 1997.

DORIGO, M.; MANIEZZO, V.; COLORNI, A. ***Ant System: Optimization by a Colony of Cooperating Agents***. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 26(1):29-41, 1996.

DORIGO, M.; STUTZLE, T. ***Ant Colony Optimization***. Cambridge. Massachusetts Institute of Technology, P.305, 2004.

EILON, S.; WATSON, C. D. T.; CHRISTOFIDES, N. ***Distribution Management: Mathematical Modelling and Practical Analysis***. London, UK: Griffin, 1971.

FELLER, W. An ***Introduction to Probability Theory and its Applications***. V.1.3.ed New York: John Wiley, 1968.

FOGLIATTI, M. C.; COSTA, N. M. **Teoria de Filas**. Editora Interciencia, 1ª Ed., p: 308, 2006.

GALVÃO, R. D.; FERREIRA FILHO, V. J. M.; RIVAS, M. P. A. **Some Computational Aspects of P-Median Type Location Problems**. Proceedings of VIII CLAIO/XXVIII Symposium of the Brazilian O.R. Society, Rio de Janeiro, Brazil, 1266-1271, 1996.

GALVÃO, R. D., CHIYOSHI, F., ESPEJO, L. G. A., RIVAS, M. P. A. **Solução do Problema de Localização de Máxima Disponibilidade Utilizando o Modelo Hipercubo**. Pesquisa Operacional, SOBRAPO 23 (1), p. 61-78, 2003.

GALVÃO R. D., CHIYOSHI F., MORABITO R.; **Towards Unified Formulations and Extensions of two Classical Probabilistic Location Models**. Computers & Operations Research 32, p. 15-33, 2005.

GALVÃO R. D.; MORABITO, R. **Emergency Service Systems: The use of the Hypercube Queuing Model in the Solution of Probabilistic Location Problems**. International Transactions in Operational Research 15, p. 1-25, 2008.

GONÇALVES M. B., NOVAES A. G., ALBINO J. C. C. **Modelos Para Localização de Serviços Emergenciais em Rodovias**. In: Simpósio Brasileiro de Pesquisa Operacional 26, Florianópolis, SC. Anais. Florianópolis, p.591-596, 1994.

GOSS, S.; ARON, S.; DENEUBOURG, J. L.; PASTEELS, J. M. **Self-Organized Shortcuts in the Argentine Ant**. Naturwissenschaften. n76, p.579-581, 1989.

HILLIER, F. S.; LIEBERMAN, G. J. **Introduction to Operations Research**. Seventh Edition, 2001.

HOFFMAN, K., WOLFE, P. **History in The Traveling Salesman Problem**. Lawler, Lenstra, Rinnooy Kan and Shmoys, eds. Wiley, 1-16, 1985.

IANNONI A. P. **Otimização da configuração e operação de sistemas médico emergenciais em rodovias utilizando o modelo hipercubo**. Tese (Doutorado em Engenharia de Produção) - Departamento de Engenharia de Produção, Universidade Federal de São Carlos, São Carlos, 2005.

IANNONI A. P., MORABITO R. **Modelo hipercubo integrado a um algoritmo genético para análise de sistemas médicos emergenciais em rodovias**. Gestão & Produção, 13(1), 93-104, 2006a.

IANNONI, A. P., MORABITO, R. **Modelo de fila hipercubo com múltiplo despacho e backup parcial para análise de sistemas de atendimento médico emergenciais em rodovias**. Pesquisa Operacional, v.26, n.3, p.493-519, 2006b.

IANNONI, A. P., MORABITO, R., SAYDAM, C. **A hypercube queueing model embedded into a genetic algorithm for ambulance deployment on highways**. Annals of Operations Research 157 (1), p. 207 – 224, 2008a.

IANNONI A. P., MORABITO, R., SAYDAM, C. ***An optimization approach for ambulance location and the districting of the response segments on highways.*** *European Journal of Operational Research*, 2008b.

IBGE- Instituto Brasileiro de Geografia e Estatística. Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=pr>>. Acesso em: 18 /03/2013.

IPPUC - Instituto de Pesquisa e Planejamento Urbano de Curitiba, <<http://www.ippuc.org.br>>, 2001.

JACOBS, L. W.; BRUSCO, M. J. ***A Local-Search Heuristic for Large Setcovering Problems.*** *Naval Research Logistics*, 42 (7), 1129-1140, 1995.

KARP, R. M. ***On the Computational Complexity of Combinatorial Problems.*** *Networks* 5, 45-68, 1975.

KENDALL, D. G. ***Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain.*** *The annals of mathematical Statistics* 24(3), 338-354, 1953.

KOIDE, R. M. ***Algoritmo de Colônia de Formigas Aplicado à Otimização de Materiais Compostos Laminados.*** Dissertação (Mestrado). Programa de Pós-graduação em Engenharia Mecânica e Materiais, Universidade Tecnológica Federal do Paraná, Curitiba, 2010.

LARSON R.C. ***Measuring the response patterns of New York city police patrol cars.*** Rand Institute. New York City, p.73, 1971.

LARSON R.C. ***Urban police patrol analysis.*** Cambridge, MIT Press, 1972

LARSON, R.C. ***Hypercube queuing model for facility location and redistricting in urban emergency services.*** *Computers and operations research* 1, 67-95,1974.

LARSON R.C.; ODONI A.R. ***Urban Operations Research.*** New Jersey, Prentice-hall, 1981.

LARSON R.C.; ODONI A.R. ***Urban Operations Research.*** 2ª ed. Dynamic Ideas, Belmont, Massachusett, 2007.

LUQUE, L. ***Análise da Aglutinação de Estados em Cadeias de Markov do Modelo hipercubo de Filas com Servidores Co-localizados.*** Dissertação (Mestrado), INPE, São José dos Campos, SP, 2007.

MARINHO, A. ***Um Estudo Sobre as Filas para Transplantes no Sistema Único de saúde Brasileira.*** *Cad. Saúde Pública*, RJ, 22(10): p 2229-2239, 2006.

MARINS, F. A., S. ***Introdução á Pesquisa Operacional.*** Cultura Acadêmica. 2012.

MAZZEO, S.; LOISEAU, I. ***An Ant Colony Algorithm for the Capacitated Vehicle Routing.*** *Eletronic Notes in Discrete Mathematics*, 18:181-186, 2004.

MENDONÇA F.C. **Aplicação do Modelo Hipercubo Baseado em Teoria de Filas, Para Análise de um Sistema Médico Emergencial em Rodovias**. Dissertação (Mestrado). Universidade Federal de São Carlos, São Carlos, 1999.

MENDONÇA F. C; MORABITO R. **Aplicação do Modelo Hipercubo para Análise de um Sistema Médico-emergencial em Rodovia**. *Gestão e Produção* 7(1), 73-91, 2000.

MORABITO, R.; MENDONÇA F. C, **Analyzing Emergency Service Ambulance Deployment on a Brazilian Highway Using the Hypercube model**. *Journal of the Operation Research Society* 52, 261-268, 2001.

MORABITO, R.; IANNONI, A. P. **Otimização da Localização das Bases de Ambulâncias e do Dimensionamento das suas Regiões de Cobertura em Rodovias**. *Produção* V 18, p.47-63, 2008.

MORAES, F. G.; SILVA, G. F.; REZENDE, T. **A Introdução à Teoria das Filas**. II Colóquio de matemática da região Centro-Oeste, 2011.

MOREIRA, D. A. **Pesquisa Operacional**. Curso introdutório. 2ª edição revista e atualizada. São Paulo. CENGAGE Learning, 2011.

MULATI, M. H.; CONSTANTINO, A. A.; SILVA, A. F. **Otimização Por Colônia de Formigas**. *Meta Heurística em Pesquisa Operacional*, pag 54-67, 2013.

NÉIA, S. S.; ARTERO, A. O.; CANTÃO, L. A. P.; CUNHA, C. B. **Roteamento de Veículos Utilizando Otimização por Colônia de Formigas e Algoritmo Genético**. *Meta Heurística em Pesquisa Operacional*, pag 219 – 236, 2013.

OLIVER, I. M.; SMITH, D. J.; HOLLAND, J. R. C. **A Study of permutation Crossover Operators on the Traveling Salesman Problem**. In: *Proceedings of the Second International Conference on Genetic Algorithms and their Applications* Hillsdale, USA. L Erlbaum Associates. Inc, p 224-230, 1987.

PEREIRA, D. A. B.; VENDRAMIN, A. C. B. K.; POHL, A. A. P. **Otimização por Colônia de Formigas e Busca Local na segmentação de Imagens**. *Anais SULCOMP*, vol.5, n.5, 2010.

PORTAL BRASIL, <http://www.brasil.gov.br/saude/2011/09/doencas-cardiovasculares>, 2011.

PRADO, D. S. **Teoria das Filas e da Simulação**. Editora: INDG, 4ª ed., v. 2, p 124, 2006.

PRESTES, A. N. **Uma Análise Experimental de Abordagens Heurísticas Aplicadas ao Problema do Caixeiro Viajante**. Dissertação (Mestrado Programa de Pós Graduação em Sistemas e Computação), Universidade Federal do Rio grande do Norte, Natal, 2006.

QUINTANA, J. F. **A Relação entre Hipertensão com Outros Fatores de Risco Para Doenças Cardiovasculares e Tratamento pela psicoterapia Cognitivo Comportamental.** Periódicos eletrônicos em psicologia. Revista da SBPH, vol 14, nº1, RJ, junho, 2011.

REVELLE, C.; HOGAN, K. ***The maximum availability location problem.*** Transportation Science, 23, 192–200, 1989.

RICHARDSON, A. J.; AMPT, E. S.; MEYBURG, A. H. ***Survey methods for transport planning.*** Parkville, Eucalyptus Press, 1995.

SILVA, E. O. A.; BENTES, C.; BAHIANSE, L.; CASTRO, M. C. S. **Uma Abordagem Paralela Baseada em Colônia de Formigas para o Problema do Caixeiro Viajante.** Cadernos do IME: Série Informática. Vol.18, 2005.

SILVA, R. M. A. **Otimização Baseada em Colônia de Formigas Aplicada ao Problema de Cobertura de Conjuntos.** Tese (Doutorado em ciência da computação), centro de informática, Universidade Federal de Pernambuco, Recife, 2003.

SILVA, L. E. **Sistema Híbrido Meta-Heurístico Baseado em Colônia de Formigas Paraconsistentes Aplicado ao Problema de Otimização em Redes Inteligentes.** Tese (Doutorado), Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal de Itajuba, Itajubá, 2012.

SIMPSON, N. C.; HANCOCK, P. G. ***Fifty Years of Operational Research and Emergency Response.*** Journal of the Operational Research Society 60, p. 126-139, 2009.

SOUZA, R. M.; MORABITO, R. **Estudos Preliminares das Flutuações de demanda do SAMU-PR.** XXVIII Encontro Nacional de Engenharia de Produção, 2008.

SOUZA, R. M. **Análise da configuração de SAMU utilizando Modelo Hipercubo de filas com prioridades e múltiplas alternativas de localização de ambulâncias.** Tese (Doutorado em engenharia de produção), Universidade Federal de São Carlos, São Carlos, 2010.

SOUZA, R. M.; MORABITO, R.; CHIYOSHII, F. Y.; IANNONI, A. P. **Análise da Configuração de SAMU Utilizando Múltiplas Alternativas de Localização de Ambulâncias.** Gestão de produção, São Carlos, v.20, n.2, p.287-302, 2013.

SWAIN, R.W. ***A Parametric Decomposition Approach for the Solution of Uncapacitated Location Problems.*** Management Science, 21, 189-198, 1971.

SWERSEY, A. J.; NGALL, E. J. ***Delivery of urban services.*** New York, Elsevier Science Publishing Company, 1986.

TAKEDA, R. A. (2000) **Uma contribuição para avaliar o desempenho de sistemas de transporte emergencial de saúde.** Tese (Doutorado em transportes), Universidade Federal de São Paulo, São Paulo, 2000.

TAKEDA R.A.; WIDMER, J. A.; MORABITO, R. **Aplicação do modelo hipercubo de filas para avaliar a descentralização de ambulâncias em um sistema urbano de atendimento médico de urgência.** *Pesquisa Operacional* 24 (1), 39-72, 2004.

TAKEDA R.A.; WIDMER, J.A.; MORABITO, R. **Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model.** *Computers & Operations Research* 34, p. 727-741, 2007.

TEITZ, M. B; BART, P. **Heuristic methods for estimating the generalized vertex median of a weighted graph.** *Gpns. Res.* 16, 955-961, 1968.

TOTH, P.; VIGO, D. **Heuristic Algorithm for the Handicapped Person Transportation Problem.** *Transportation Science*, Vol. 31, n.1, p. 60-71, 1997.

WEYLAND, D.; SALANI, M.; MONTEMANNI, R.; GAMBARDELLA, L. M. **Vehicle Routing for Exhausted Oil Collection.** *Journal of Traffic and logistics Engineering*. Vol1, nº.1 March, 2013.

WOLFF, R. W. **Stochastic Modeling and Theory of Queues.** Nova Jersey, prentice Hall, 1989.

YILDRIM, U. M.; ÇATAY, B. **An Ant Colony Algorithm for Time-Dependent Vehicle Routing problem with Time Windows.** *Operations Research proceedings, V.2008, pp337-342* Sabanci University, Orhanh, Tuzla, 34956 Istanbul, Turkey, 2009.

ANEXOS

ANEXO 1 - EXEMPLO ILUSTRATIVO DO ALGORITMO BASEADO NO ACO.....	163
ANEXO 2 - DISTRIBUIÇÃO DE POISSON E EXPONENCIAL.....	174

$$\eta_{ij} = \frac{1}{d_{ij}}, \text{ onde } n \text{ é a matriz de atratividade e } d_{ij} \text{ é a matriz de distâncias.}$$

TABELA 3 - MATRIZ DE ATRATIVIDADE N

-	1	0,5	0,33	0,25	0,2	0,33	0,25
1	-	1	0,5	0,33	0,25	0,5	0,33
0,5	1	-	1	0,5	0,33	1	0,5
0,33	0,5	1	-	1	0,5	0,5	0,33
0,25	0,33	0,5	1	-	1	0,33	0,25
0,2	0,25	0,33	0,5	1	-	0,25	0,2
0,33	0,5	1	0,5	0,33	0,25	-	1
0,25	0,33	0,5	0,33	0,25	0,2	1	-

1º Passo: Definir a saída (cidades) das formigas, cada duas formigas criam duas rotas. Conforme podemos ver na (TABELA 4), as formigas 1 e 2 saíram das cidades 1 e 4, de maneira análoga as formigas 3 e 4, saíram das cidades 4 e 2, e as formigas 5 e 6 saíram das cidades 2 e 7.

Saída aleatória das formigas:

TABELA 4 - CIDADE ALEATÓRIA

Formiga	Cidade
1	1
2	4
3	4
4	2
5	2
6	7

1ª Iteração:

Calculo da matriz de probabilidade:

$$K_{ij} = \begin{cases} i = \text{rotas 1 e 2} \\ j = \text{formigas 1, 2, 3, 4, 5 e 6} \end{cases}$$

$$P_{ij}^k = \frac{(T_{ij})^\alpha \cdot (\eta_{ij})^\beta}{\sum_{K \in N_i^k} (T_{ij})^\alpha \cdot (\eta_{ij})^\beta} \quad (1)$$

Onde: $\alpha = 1$
 $\beta = 2$

T_{ij} Refere-se à quantidade de feromônio associada à aresta i, j.

η_{ij} Refere-se à atratividade associada à aresta i, j.

Formiga 1: saindo aleatoriamente da cidade 1 começa sua rota na cidade 1.

Restrições: não poderá passar nas cidades 1 e 4 (pois a formiga 2 saiu dessa cidades).

$$P_{12}^1 = \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,20)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2} = \frac{1}{1,52} = 0,66$$

= 66 %

$$P_{13}^1 = \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,20)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2} = 0,16 = 16 \%$$

$$P_{15}^1 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,20)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2} = 0,041 = 4,1 \%$$

$$P_{16}^1 = \frac{(1)^1 \cdot (0,2)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,20)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2} = 0,026 = 2,6 \%$$

$$P_{17}^1 = \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,20)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2} = 0,0712 = 7 \%$$

$$P_{18}^1 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,20)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2} = 0,041 = 4,1 \%$$

Mesma restrição para a formiga 2, saindo da cidade 4 não poderá passar nas cidades 4 e 1 (formiga 1 saiu da cidade 1). Logo as probabilidades serão calculadas com essas exceções.

$$P_{42}^2 = \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2} = 0,087 = 8,7 \%$$

$$P_{43}^2 = \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2} = 0,35 = 35 \%$$

$$P_{45}^2 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2} = 0,35 = 35 \%$$

$$P_{46}^2 = \frac{(1)^1 \cdot (0,2)^2}{(1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2} = 0,087 = 8,7 \%$$

$$P_{47}^2 = \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2} = 0,087 = 8,7 \%$$

$$P_{48}^2 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2} = 0,038 = 3,8 \%$$

Raciocínio análogo para as formigas 3 e 4, saindo das cidades 2 e 7 respectivamente, não poderão passar por essas cidades. Novamente calculam-se as probabilidades, contudo, acrescentando essas restrições.

$$P_{21}^3 = \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2} = 0,395 = 39,5 \%$$

$$P_{23}^3 = \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2} = 0,395 = 39,5 \%$$

$$P_{24}^3 = \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2} = 0,0988 = 9,88 \%$$

$$P_{25}^3 = \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2} = 0,043 = 4,3 \%$$

$$P_{26}^3 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2} = 0,024 = 2,4 \%$$

$$P_{28}^3 = \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (1)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2} = 0,043 = 4,3 \%$$

$$\begin{aligned}
P_{71}^4 &= \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (0,33)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (1)^2} = 0,043 = 4,3 \% \\
P_{73}^4 &= \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (0,33)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (1)^2} = 0,395 = 39,5 \% \\
P_{74}^4 &= \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (0,33)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (1)^2} = 0,0988 = 9,88 \% \\
P_{75}^4 &= \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (0,33)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (1)^2} = 0,043 = 4,3 \% \\
P_{76}^4 &= \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,33)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (1)^2} = 0,024 = 2,4 \% \\
P_{78}^4 &= \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (0,33)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (1)^2} = 0,395 = 39,5 \%
\end{aligned}$$

Da mesma forma para o último par de formigas, 5 e 6. Saindo das cidades 6 e 8 respectivamente, não poderão passar pelas cidades 2 e 7. Calculam-se as probabilidades através da equação 1.1, contudo acrescentado essas restrições.

$$\begin{aligned}
P_{61}^5 &= \frac{(1)^1 \cdot (0,2)^2}{(1)^1 \cdot (0,2)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,25)^2} = 0,026 = 2,6 \% \\
P_{62}^5 &= \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,2)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,25)^2} = 0,042 = 4,1 \% \\
P_{63}^5 &= \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (0,2)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,25)^2} = 0,071 = 7,1 \% \\
P_{64}^5 &= \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (0,2)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,25)^2} = 0,164 = 16,4 \% \\
P_{65}^5 &= \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (0,2)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,25)^2} = 0,65 = 65 \% \\
P_{67}^5 &= \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,2)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (1)^2 + (1)^1 \cdot (0,25)^2} = 0,041 = 4,1 \%
\end{aligned}$$

$$P_{81}^6 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,1)^2} = 0,0392 = 3,9 \%$$

$$P_{82}^6 = \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,1)^2} = 0,068 = 6,8 \%$$

$$P_{83}^6 = \frac{(1)^1 \cdot (0,5)^2}{(1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,1)^2} = 0,1569 = 15,69 \%$$

$$P_{84}^6 = \frac{(1)^1 \cdot (0,33)^2}{(1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,1)^2} = 0,068 = 6,8 \%$$

$$P_{85}^6 = \frac{(1)^1 \cdot (0,25)^2}{(1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,1)^2} = 0,0392 = 3,9 \%$$

$$P_{87}^6 = \frac{(1)^1 \cdot (1)^2}{(1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,5)^2 + (1)^1 \cdot (0,33)^2 + (1)^1 \cdot (0,25)^2 + (1)^1 \cdot (0,1)^2} = 0,627 = 62,7 \%$$

A TABELA 5 apresenta a matriz de probabilidade:

TABELA 5 - MATRIZ DE PROBABILIDADE P

K11	0	0,61	0,15	0	0,038	0,0244	0,066	0,038
K21	0	0,087	0,35	0	0,35	0,087	0,087	0,038
K12	0,395	0	0,395	0,0988	0,043	0,024	0	0,043
K22	0,043	0	0,395	0,0988	0,043	0,0247	0	0,395
K13	0,026	0,041	0,071	0,164	0,65	0	0,041	0
K23	0,0392	0,068	0,1569	0,068	0,0392	0	0,62	0

Encontrar o maior valor das probabilidades para cada par de formigas, em caso de empate na mesma posição, escolhe-se a segunda maior, procurando pegar pontos diferentes. Em seguida, zeram-se as posições que a formiga já passou.

TABELA 6 - ROTA EM CONSTRUÇÃO

Formiga	Rota	
1- K11	1	2
2- K21	4	3
3- K12	2	1
4- K22	7	3
5- K13	6	5
6- K23	8	7

2º Passo: Calcular as novas probabilidades acrescentando as restrições

Formiga 1: saindo da cidade 2 (não passa pelas cidades 1, 2, 3 e 4. calcula as probabilidades;

Formiga 2: saindo da cidade 3 (não passa pelas cidades 1, 2, 3 e 4. calcula as probabilidades;

Formiga 3: saindo da cidade 1 (não passa pelas cidades 1, 2, 7 e 3. calcula as probabilidades;

Formiga 4: saindo da cidade 3 (não passa pelas cidades 1, 2, 7 e 3. calcula as probabilidades;

Formiga 3: saindo da cidade 5 (não passa pelas cidades 5, 6, 7 e 8. Calcula as probabilidades;

Formiga 3: saindo da cidade 7 (não passa pelas cidades 5, 6, 7 e 8. Calcula as probabilidades;

Calcula-se nova matriz de probabilidade de maneira análoga, conforme TABELA 7:

TABELA 7 - MATRIZ DE PROBABILIDADE P

K11	0	0	0	0	0,205	0,1178	0,471	0,205
K21	0	0	0	0	0,155	0,067	0,621	0,155
K12	0	0	0	0,39	0,23	0,15	0	0,23
K22	0	0	0	0,62	0,15	0,068	0	0,15
K13	0,044	0,076	0,017	0,704	0	0	0	0
K23	0,067	0,155	0,62	0,155	0	0	0	0

TABELA 8 – ROTA EM CONSTRUÇÃO

Formiga	Rota		
K11	1	2	7
K21	4	3	5
K12	2	1	4
K22	7	3	5
K13	6	5	4
K23	8	7	3

3º Passo:

Formiga 1: saindo da cidade 7 (não passa pelas cidades 1, 2, 3, 4, 5 e 7. calcula as probabilidades;

Formiga 1: saindo da cidade 5 (não passa pelas cidades 1, 2, 3, 4, 5 e 7. calcula as probabilidades;

Formiga 2: saindo da cidade 4 (não passa pelas cidades 1, 2, 7, 3, 4 e 5. calcula as probabilidades;

Formiga 2: saindo da cidade 5 (não passa pelas cidades 1, 2, 7, 3, 4 e 5. calcula as probabilidades;

Formiga 3: saindo da cidade 4 (não passa pelas cidades 3, 4, 5, 6, 7 e 8. calcula as probabilidades;

Formiga 3: saindo da cidade 3 (não passa pelas cidades 3, 4, 5, 6, 7 e 8. calcula as probabilidades;

TABELA 9 - MATRIZ DE PROBABILIDADE P

K11	0	0	0	0	0	0,0588	0	0,94
K21	0	0	0	0	0	0,94	0	0,058
K12	0	0	0	0	0	0,69	0	0,30
K22	0	0	0	0	0	0,94	0	0,05
K13	0,3	0,69	0	0	0	0	0	0
K23	0,2	0,8	0	0	0	0	0	0

TABELA 10 - ROTA EM CONSTRUÇÃO

Formiga	Rota				
K11	1	2	7	8	1
K21	4	3	5	6	4
K12	2	1	4	6	2
K22	7	3	5	8	7
K13	6	5	4	1	6
K23	8	7	3	2	8

Calcular os LKs, conforme TABELA 11.

TABELA 11 - CUSTOS LKs

Formiga	Rota					LK	LK total
1- K11	1	2	7	8	1	8	14
2- K21	4	3	5	6	4	6	
3- K12	2	1	4	6	2	10	18
4- K22	7	3	5	8	7	8	
5- K13	6	5	4	1	6	10	16
6- K23	8	7	3	2	8	6	

Final da 1ª iteração. (Melhor rota, menor L é a da formiga 1 e 2)

2ª iteração:

Atualizar a matriz de feromônio:

$$T = T_{ij} = (1 - \rho)T_{ij} + \sum_{k=1}^m \Delta T_{ij}^k$$

$$\Delta T_{ij} = \frac{1}{L_k}$$

Ex: Na aresta 1-2: as formigas que depositaram feromônio nesta aresta são: 1 e 2.

Logo;

- ▶ $T_{12} = (1-0,5).T_{12} + (1/14+1/18)$
- ▶ $T_{12} = 0,5.1 + (1/14+1/18)=0,627$
- ▶ $T_{12} = T_{21}$

1- ρ = evaporação do feromônio

- ▶ Aresta 1- 3: Formiga 2
- ▶ $T_{13} = T_{31}$
- ▶ $T_{13} = (1-0,5)= 0,50$
- ▶ Aresta 1- 4: Formiga 3

- ▶ $T_{14} = T_{41}$
- ▶ $T_{14} = (1-0,5)+(1/16+ 1/18) = \mathbf{0,62}$

De maneira análoga os de mais cálculos.

Nova matriz de feromônio, TABELA 12:

TABELA 12 – NOVA MATRIZ DE FEROMONIO

1	0,627	0,50	0, 62	0,5	0,563	0,5	0,571
0,627	1	0,563	0,5	0,5	0,555	0,57	0,563
0,50	0,563	1	0,57	0,63	0,50	0,62	0,5
0,62	0,5	0,57	1	0,56	0,57	0,5	0,50
0,5	0,5	0,63	0,56	1	0,634	0,50	0,55
0,563	0,555	0,50	0,57	0,634	1	0,5	0,5
0,5	0,57	0,62	0,5	0,50	0,5	1	0,689
0,571	0,563	0,5	0,50	0,55	0,5	0,689	1

E assim continua o processo até que algum critério de parada seja satisfeito. Ao final têm-se três conjuntos de pontos com dois átomos geográficos cada. Escolhe-se o de menor custo, conseqüentemente melhores divisões.

ANEXO 2

DISTRIBUIÇÃO DE POISSON E EXPONENCIAL

O objetivo da segunda hipótese do modelo hipercubo de filas é provar que as chegadas em cada átomo constituem um processo de Poisson, também chamado de processo de contagem de incrementos, onde a probabilidade de ocorrer um evento em um pequeno intervalo de tempo é proporcional a λ , e a probabilidade de ocorrer dois ou mais eventos é igual a 0.

Facilmente pode-se demonstrar que: “Se o número de chegadas em um determinado intervalo de tempo tem distribuição de Poisson de parâmetro $\lambda > 0$, os intervalos entre as chegadas sucessivas ao sistema têm distribuição exponencial de parâmetro $1/\lambda$ ”. A recíproca é verdadeira, e a demonstração completa encontra-se em Wolff (1989).

Nesse estudo de caso, constatou-se que as chegadas dos chamados em todos os átomos do SAE móvel, constituem processos de contagem com incrementos independentes. Logo, para provar que o número de chamadas em cada átomo segue o padrão de Poisson, basta determinar as taxas médias de solicitação pelo sistema durante o período de observação, bem como, os intervalos médios entre as chegadas sucessivas em todos os átomos.

Vale lembrar que o número de chamadas durante o período de pico, considerando cada átomo individualmente é muito pequeno. Fato que dificultou consideravelmente a aplicação das ferramentas estatística para validade tal hipótese.

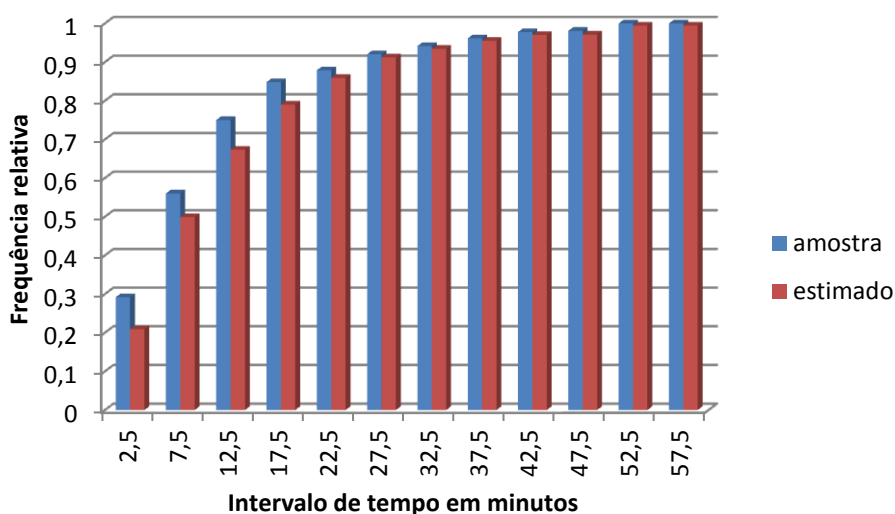
Contudo, os chamados ocorrem independentemente em cada átomo, e o número de chamados em cada átomo representa uma proporção do total do sistema, conforme apresentado na seção 4.4.2.1. Com isso, foram determinadas as probabilidades p_j que reproduzem as possibilidades de ocorrência de um chamado qualquer no sistema ter sido originado no átomo j .

A fim de aplicar a metodologia proposta para verificar a distribuição estatística, foi calculada a diferença entre o horário de chegada de cada chamado. Isso foi realizado para todos os chamados para os sete dias da amostra. Feito isso

os dados foram agrupados em uma única coluna. O teste foi realizado conforme mencionado anteriormente, no aplicativo do *software Arena*.

Foi realizado um teste de aderência com o intuito de verificar estatisticamente a distribuição dos chamados entre as chegadas sucessivas. Foi aplicado o Método Qui-Quadrado e *Kolmogorov-Smirnov*, ao nível de significância de $\alpha = 0,15$.

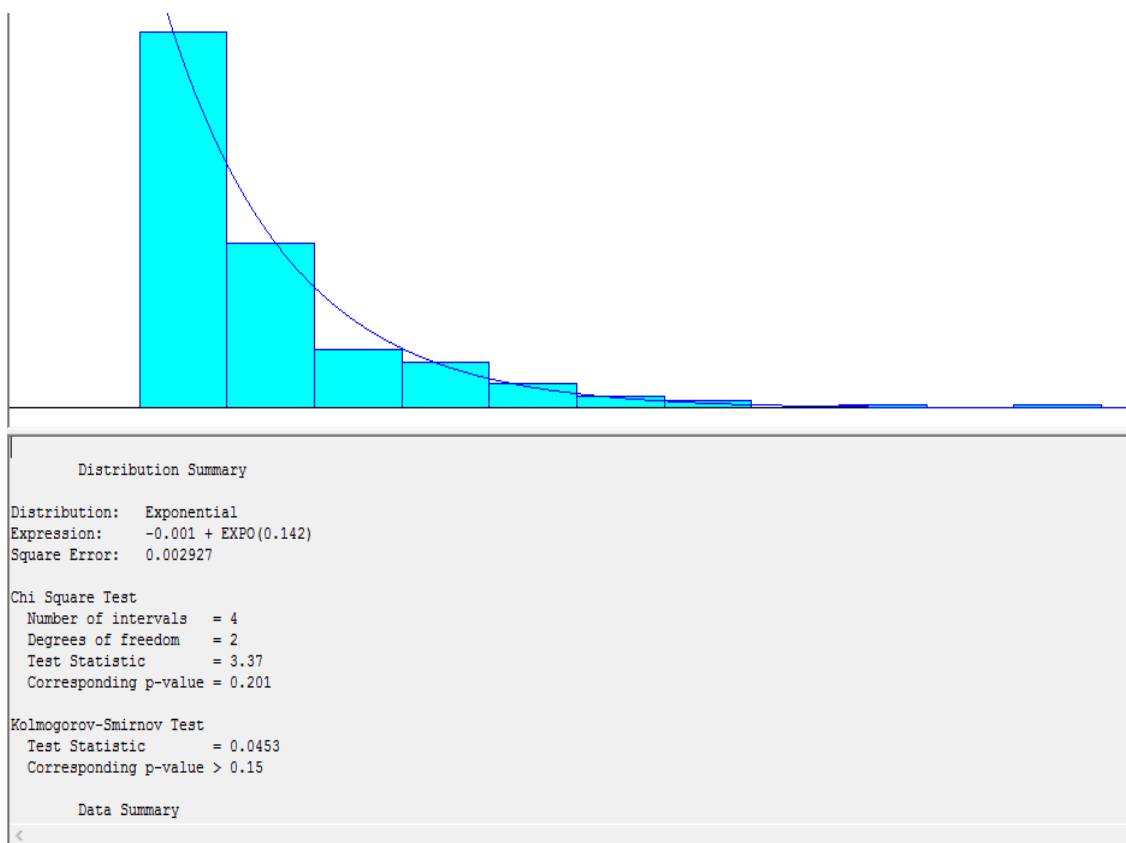
A FIGURA abaixo mostra os dados da amostra e os valores obtidos através da aplicação do modelo exponencial para os sete dias da amostra.



A diferença máxima entre as frequências relativas acumuladas na amostra e estimadas pela distribuição exponencial foi de 0,0828. O valor tabelado pelo Método *Kolmogorov-Smirnov* foi de 0,0956. O valor crítico obtido pelo método é maior que o valor máximo obtido pela amostra, logo, não se pode rejeitar a hipótese de que os intervalos entre as chegadas sucessivas ao SAE móvel tem distribuição exponencial, logo, o número de chamadas por intervalo de tempo tem distribuição de Poisson. A TABELA a seguir apresenta: $G(x)$ - frequências acumuladas observadas; $F(x_i)$ - função de distribuição de probabilidade exponencial.

Intervalo (minutos)	x_i - média dos intervalos	Quantidade de observações	$G(x)$	$F(x_i)$	$G(x) - F(x_i)$
0; 5	2,5	60	0,2918	0,209	0,0828
5; 10	7,5	56	0,5604	0,499	0,0614
10; 15	12,5	42	0,7502	0,6733	0,0769
15; 20	17,5	15	0,8481	0,79	0,0581
20; 25	22,5	11	0,8788	0,859	0,0198
25; 30	27,5	6	0,9203	0,9122	0,0081
30; 35	32,5	5	0,9416	0,9344	0,0072
35; 40	37,5	4	0,9614	0,955	0,0064
40; 45	42,5	4	0,9777	0,96999	0,00771
45; 50	47,5	2	0,9809	0,971	0,0099
50; 55	52,5	2	0,9999	0,9944	0,0055
55; 60	57,5	1	1	0,99444	0,00556
Total		208 - chamados			

Testes de Qui-quadrado e Kolmogorov-Smirnov para aderência da amostra a uma distribuição Exponencial mostrada na figura abaixo:



Conclui-se que não se pode rejeitar a hipótese de que os intervalos entre as chegadas sucessivas ao SAE móvel tem distribuição exponencial, logo, o número de chegadas por intervalo de tempo tem distribuição de Poisson.